

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:55:55 ; Search time 16.351 Seconds
(without alignments)
4244.936 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 722

Sequence: 1 MGQFMSVFRIMNTATVLA.....NTFNPGSERTWNLRYSF 722

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	172	23.8	708	2 B81038	Tomb-dependent rec
2	11	1.5	853	2 AC2079	ferrichrome-iron r
3	9	1.2	629	2 C83997	1-deoxyxylose-5-
4	9	1.2	725	2 D81976	probable ferric si
5	9	1.2	725	2 H81030	Tomb-dependent rec
6	8	1.1	284	2 T39813	hypothetical prote
7	8	1.1	532	1 WZBEF5	59k transcription
8	8	1.1	644	2 G70794	probable alpha-iso
9	8	1.1	696	2 AC0547	ferritoxamine B rec
10	8	1.1	720	2 A36942	Fe(III)-pyochelin
11	8	1.1	753	2 AE0187	probable iron-side
12	8	1.1	775	2 T30917	hypothetical prote
13	8	1.1	813	2 A40601	ferrityoverdine re
14	8	1.1	815	2 H83345	ferrityoverdine re
15	8	1.1	837	2 B87759	protein D1007.15 l
16	8	1.1	1274	2 AD2109	hypothetical prote
17	8	1.1	78	2 AD2109	hypothetical prote
18	7	1.0	84	2 AC1886	hypothetical prote
19	7	1.0	102	2 C81660	thioredoxin TC0826
20	7	1.0	108	2 C97545	hypothetical prote
21	7	1.0	108	2 AD2764	conserved hypotet
22	7	1.0	112	2 F87193	probable secreted
23	7	1.0	123	2 S29304	hypothetical prote
24	7	1.0	127	2 T38756	60s ribosomal prot
25	7	1.0	127	2 T39562	60s ribosomal prot
26	7	1.0	130	2 A84120	ATP synthase prote
27	7	1.0	146	2 S71256	ribosomal protein
28	7	1.0	148	2 C86732	transcription regu
29	7	1.0	149	1 R8BY29	ribosomal protein

30	7	1.0	155	2 S77384	hypothetical prote
31	7	1.0	166	2 AH3359	probable competent
32	7	1.0	165	2 T10313	hypothetical prote
33	7	1.0	173	2 F83062	hypothetical prote
34	7	1.0	174	2 C86231	hypothetical prote
35	7	1.0	177	1 PWECD	H+-transporting tw
36	7	1.0	177	2 AC0954	ATP synthase delta
37	7	1.0	177	2 F86059	H+-transporting two
38	7	1.0	177	2 AH0500	H+-transporting tw
39	7	1.0	177	2 E91213	H+-transporting tw
40	7	1.0	182	2 D82490	hypothetical prote
41	7	1.0	191	2 T04031	hypothetical prote
42	7	1.0	201	2 E95072	nitroreductase fam
43	7	1.0	201	2 B97940	nitroreductase (BC
44	7	1.0	208	2 B86874	hypothetical prote
45	7	1.0	209	2 AC2841	outer membrane pro
46	7	1.0	211	2 S47347	outer membrane pro
47	7	1.0	214	2 T49351	endothelin 3 precu
48	7	1.0	217	2 D82407	outer membrane pro
49	7	1.0	219	2 H97618	25k outer-membrane
50	7	1.0	220	2 AE2360	hypothetical prote
51	7	1.0	221	2 T48101	hypothetical prote
52	7	1.0	236	2 G75324	hypothetical prote
53	7	1.0	241	2 T51206	hypothetical prote
54	7	1.0	241	2 S39718	spore coat polysac
55	7	1.0	260	2 AD2584	hypothetical prote
56	7	1.0	260	2 D97366	firca protein (AF19
57	7	1.0	268	2 A70379	conserved hypotet
58	7	1.0	269	2 D84312	sulfate transport
59	7	1.0	275	2 H87538	hypothetical prote
60	7	1.0	280	2 T03942	rRNA N-glycosidase
61	7	1.0	278	2 F75057	hypothetical prote
62	7	1.0	286	2 H64113	hemagglutinin ha91
63	7	1.0	295	2 S76136	hypothetical prote
64	7	1.0	306	2 B85436	RNA-binding like p
65	7	1.0	307	2 AR3487	dieneacetone hydro
66	7	1.0	310	2 T31799	hypothetical prote
67	7	1.0	312	2 A64660	site-specific DNA-
68	7	1.0	318	2 G90707	hypothetical prote
69	7	1.0	318	2 C85558	hypothetical prote
70	7	1.0	318	2 AD0575	ferriterobactin-
71	7	1.0	318	2 JV0045	ferriterobactin-
72	7	1.0	319	2 F69483	conserved hypotet
73	7	1.0	321	2 B71854	type II DNA modifi
74	7	1.0	324	2 C95048	3-oxoacyl-(acyl-ca
75	7	1.0	324	2 C95048	3-oxoacyl-(acyl-ca
76	7	1.0	329	2 T49953	pxr10 peroxidase-1
77	7	1.0	330	2 A70422	hypothetical prote
78	7	1.0	331	2 G69830	lipoate-protein 11
79	7	1.0	334	2 H64079	ornithine carbamoy
80	7	1.0	339	2 T25158	hypothetical prote
81	7	1.0	339	2 AE2212	hypothetical prote
82	7	1.0	342	2 AH1071	conserved hypotet
83	7	1.0	344	2 T11620	hypothetical prote
84	7	1.0	351	2 S25520	outer membrane por
85	7	1.0	354	2 C71368	probable UDP-N-ace
86	7	1.0	360	2 C82220	hypothetical prote
87	7	1.0	363	2 F43680	D363 protein - Af
88	7	1.0	372	2 G83707	hypothetical prote
89	7	1.0	373	1 VGBEKA	glycoprotein gX pr
90	7	1.0	382	1 A60112	42k membrane antiq
91	7	1.0	382	2 S15578	ipac protein - Shi
92	7	1.0	382	2 T04347	teosine branched l
93	7	1.0	383	2 T34603	xyranase A - Strep
94	7	1.0	394	2 F95973	probable sugar upt
95	7	1.0	395	2 A43700	BNS1 protein - hum
96	7	1.0	409	2 T47298	probable replicati
97	7	1.0	411	1 VGBEG2	glycoprotein G pre
98	7	1.0	411	2 T01388	oxidoreductase hom
99	7	1.0	412	2 D72467	hypothetical prote
100	7	1.0	421	2 T31787	hypothetical prote
101	7	1.0	421	2 E71363	hypothetical prote
102	7	1.0	423	2 A70947	hypothetical prote

103	7	1.0	426	2	E87321	176	7	1.0	1280	2	G96796	hypothetical prote
104	7	1.0	437	2	C64113	177	7	1.0	1283	2	T13799	neurexin IV - frui
105	7	1.0	443	2	S57328	178	7	1.0	1341	2	JG0166	LaMDRI protein - L
106	7	1.0	454	2	E70448	179	7	1.0	1388	2	T38720	chromodomain helic
107	7	1.0	454	2	T12539	180	7	1.0	1514	2	S70099	hypothetical prote
108	7	1.0	456	2	T11665	181	7	1.0	1545	2	T14288	hypothetical prote
109	7	1.0	457	2	H84716	182	7	1.0	1557	2	T29132	DNA (cytosine-5')-
110	7	1.0	461	2	T00918	183	7	1.0	1599	2	T15854	hypothetical prote
111	7	1.0	477	1	JS0589	184	7	1.0	1616	2	T47801	hypothetical prote
112	7	1.0	477	2	S55494	185	7	1.0	1643	1	RRWGNV	genome polyprotein
113	7	1.0	478	2	T05157	186	7	1.0	1761	2	T14289	DNA (cytosine-5')-
114	7	1.0	479	2	C97508	187	7	1.0	1772	2	T36105	probable large gly
115	7	1.0	479	2	AG2726	188	7	1.0	2014	2	S46622	probable membrane
116	7	1.0	485	2	S48650	189	7	1.0	2164	1	GNNY89	genome polyprotein
117	7	1.0	495	2	T01230	190	7	1.0	2551	2	B98047	hypothetical prote
118	7	1.0	509	1	D70703	191	7	1.0	4572	2	S57908	hypothetical 527K
119	7	1.0	505	2	T51741	192	6	0.8	35	2	H86021	hypothetical prote
120	7	1.0	506	2	T14907	193	6	0.8	39	2	AB3109	hypothetical prote
121	7	1.0	506	2	AE0596	194	6	0.8	44	2	D71666	ribosomal protein
122	7	1.0	506	2	S73862	195	6	0.8	44	2	S36977	hypothetical prote
123	7	1.0	514	2	G81935	196	6	0.8	44	2	H97816	50S ribosomal prot
124	7	1.0	514	2	A81170	197	6	0.8	50	2	G91003	hoat-nuclease inhi
125	7	1.0	518	2	F70687	198	6	0.8	53	2	A82862	hypothetical prote
126	7	1.0	520	2	E87621	199	6	0.8	55	2	A37238	autoimmune epitope
127	7	1.0	534	2	T30019	200	6	0.8	57	2	G98090	hypothetical prote
128	7	1.0	550	2	E80723	201	6	0.8	59	2	D84234	hypothetical prote
129	7	1.0	554	2	E85574	202	6	0.8	60	2	S62076	M-like protein pre
130	7	1.0	554	2	S37936	203	6	0.8	60	2	D82277	hypothetical prote
131	7	1.0	558	1	S68203	204	6	0.8	62	2	S60788	M protein precurs
132	7	1.0	564	2	S76672	205	6	0.8	63	2	F69114	hypothetical prote
133	7	1.0	564	2	T00135	206	6	0.8	63	2	S44634	thiG protein AGR C
134	7	1.0	574	2	B29677	207	6	0.8	65	2	B97667	thiamin biosynthes
135	7	1.0	577	2	E72396	208	6	0.8	65	2	AF2891	hypothetical immun
136	7	1.0	582	2	T51625	209	6	0.8	65	4	IMBP14	hypothetical prote
137	7	1.0	583	2	B85063	210	6	0.8	67	2	G82702	hypothetical prote
138	7	1.0	585	2	C82955	211	6	0.8	67	2	H97845	hypothetical prote
139	7	1.0	608	2	G96575	212	6	0.8	67	2	A13396	hypothetical prote
140	7	1.0	632	2	T18692	213	6	0.8	68	2	S60797	M protein precurs
141	7	1.0	635	2	C81861	214	6	0.8	69	2	D75517	hypothetical prote
142	7	1.0	661	2	S19646	215	6	0.8	70	1	R5EC31	ribosomal protein
143	7	1.0	702	2	T23651	216	6	0.8	70	2	G86083	50S ribosomal subu
144	7	1.0	711	2	S88443	217	6	0.8	70	2	G91236	50S ribosomal subu
145	7	1.0	711	2	C70606	218	6	0.8	70	2	T41875	P10 orf137 - Bomby
146	7	1.0	714	2	S68893	219	6	0.8	73	2	R82344	rft-related prote
147	7	1.0	716	1	WZBEE6	220	6	0.8	75	2	S62073	M-like protein pre
148	7	1.0	731	2	A83410	221	6	0.8	76	2	S60810	M protein precurs
149	7	1.0	747	2	B47093	222	6	0.8	76	2	S61818	M-like protein ann
150	7	1.0	759	1	C64345	223	6	0.8	78	2	D34509	ig light chain C r
151	7	1.0	771	2	AF2056	224	6	0.8	78	2	A97974	degenerate transpo
152	7	1.0	778	2	G98221	225	6	0.8	79	2	S62072	protein M-like pro
153	7	1.0	778	2	A13064	226	6	0.8	79	2	S61802	M-like protein ann
154	7	1.0	778	2	T43223	227	6	0.8	79	2	S62078	M-like protein pre
155	7	1.0	778	2	T45221	228	6	0.8	80	2	S61801	M-like protein ann
156	7	1.0	809	2	S28899	229	6	0.8	81	2	S61807	M-like protein ann
157	7	1.0	811	2	T39998	230	6	0.8	81	2	G98177	hypothetical prote
158	7	1.0	819	2	S15169	231	6	0.8	82	2	S61816	M-like protein ann
159	7	1.0	823	2	S48911	232	6	0.8	82	2	B90706	hypothetical prote
160	7	1.0	863	2	S74447	233	6	0.8	82	2	S29722	cellular adhesion
161	7	1.0	869	2	A82055	234	6	0.8	82	2	B82862	hypothetical prote
162	7	1.0	875	2	T12794	235	6	0.8	82	2	E85556	hypothetical prote
163	7	1.0	960	1	A39651	236	6	0.8	82	2	B64791	ybdJ protein - Esc
164	7	1.0	963	2	T09478	237	6	0.8	83	2	S61086	M protein precurs
165	7	1.0	985	2	S67255	238	6	0.8	83	2	S61824	M-like protein ann
166	7	1.0	987	2	T40241	239	6	0.8	83	2	S62075	M-like protein ann
167	7	1.0	1029	2	F86210	240	6	0.8	84	2	C90447	partial transposas
168	7	1.0	1042	2	T13647	241	6	0.8	85	2	S62083	M-like protein ann
169	7	1.0	1057	2	H83273	242	6	0.8	85	2	S14026	hypothetical prote
170	7	1.0	1068	2	E81965	243	6	0.8	86	2	S62077	M-like protein ann
171	7	1.0	1082	2	H81020	244	6	0.8	87	2	T15407	hypothetical prote
172	7	1.0	1158	2	A83285	245	6	0.8	88	2	S60795	M protein precurs
173	7	1.0	1174	2	C97686	246	6	0.8	88	2	S60812	M protein precurs
174	7	1.0	1174	2	A82911	247	6	0.8	89	2	S61821	M-like protein ann
175	7	1.0	1239	2	I49705	248	6	0.8	90	2	S62074	M-like protein ann

succinylarginine d
tetrahydrofolylpol
uvrH protein - Eme
signal recognition
hypothetical prote
translation elonga
hypothetical prote
hypothetical prote
endo-1,4-beta-xyla
C3HC4 type zinc fi
NADH dehydrogenase
blue-light photore
DNA photolyase [im
catalase (EC 1.11.
probable RNA helic
tldD homolog Rv231
RNA helicase RH8 [I
trans-cinnamate 4-
histidine ammonia-
dihydrofolate reduc
amidophosphoribosy
amidophosphoribosy
probable gabbl pro
hypothetical prote
hypothetical prote
probable fumarate
probable fumarate
regulatory protein
tyrosine N-monooxy
hypothetical prote
hypothetical prote
complement C9 prec
ABC transporter, A
MAP3K alpha protei
hypothetical prote
probable sodium/pr
probable MEK kinas
hypothetical prote
hypothetical prote
DNA repair protein
hypothetical prote
double-stranded RN
probable fadE34 pr
hypothetical prote
77.8K DNA helicase
probable oxidoredu
cellulase (EC 3.2.
replication licens
cation transportin
glucose dehydrogen
glucose dehydrogen
hypothetical prote
DNA damage checkpo
ferric-pseudobacti
hypothetical prote
ferric-pseudobacti
hypothetical prote
ferrichrome-iron r
extracellular nucl
hypothetical yomG
discs-large tumor
ubiquitin thiolest
REV1 protein - yea
probable guanine n
hypothetical prote
hypothetical prote
ribonuclease E PA2
serotype-1-specifi
pyruvate carboxyla
pyruvate carboxyla
pyruvate carboxyla
glutamate receptor

249	6	0.8	90	2	S62081	M-like protein pre	322	6	0.8	131	2	A12329	30S ribosomal prot
250	6	0.8	90	2	B64031	hypothetical prote	323	6	0.8	131	2	C65089	hypothetical prote
251	6	0.8	90	2	D81684	conserved hypotet	324	6	0.8	132	2	AD0108	conserved hypotet
252	6	0.8	91	2	G90824	hypothetical prote	325	6	0.8	132	2	G72462	hypothetical prote
253	6	0.8	92	2	S61809	M-like protein em	326	6	0.8	132	2	A13314	acyl-CoA hydrolase
254	6	0.8	92	2	A84637	hypothetical prote	327	6	0.8	132	2	T50779	copper chaperone h
255	6	0.8	93	2	D69960	hypothetical prote	328	6	0.8	133	2	F64700	biopolymer transpo
256	6	0.8	94	2	S60858	M protein precuro	329	6	0.8	133	2	B82550	conserved hypotet
257	6	0.8	94	2	S61805	M-like protein em	330	6	0.8	133	2	T14238	probable secreted
258	6	0.8	94	2	S62084	M-like protein em	331	6	0.8	133	2	AE1971	hypothetical prote
259	6	0.8	94	2	B72867	fibrous body prote	332	6	0.8	134	2	T25527	hypothetical prote
260	6	0.8	96	2	D47624	Ig heavy chain V-I	333	6	0.8	134	2	B97642	hypothetical prote
261	6	0.8	96	2	S61827	M-like protein em	334	6	0.8	134	2	AE2866	hypothetical prote
262	6	0.8	96	2	S71519	M protein type 19	335	6	0.8	136	2	I49013	thymic shared anti
263	6	0.8	96	2	S10069	regulatory protein	336	6	0.8	137	2	P90565	30S ribosomal prot
264	6	0.8	97	2	S61811	M-like protein em	337	6	0.8	137	2	S34620	probable cytochrom
265	6	0.8	97	2	AE3595	hypothetical cytos	338	6	0.8	137	2	B75447	hypothetical prote
266	6	0.8	98	2	S61079	M protein precuro	339	6	0.8	137	2	A97256	uncharacterized pr
267	6	0.8	98	2	AB2100	hypothetical prote	340	6	0.8	138	1	R3NT11	ribosomal protein
268	6	0.8	99	2	D69060	hypothetical prote	341	6	0.8	138	1	R3PM11	ribosomal protein
269	6	0.8	101	2	C72502	p12 protein - beet	342	6	0.8	138	1	OCBPL	host-nuclease inh
270	6	0.8	101	2	B44503	probable thioredox	343	6	0.8	138	2	B35525	ribosomal protein
271	6	0.8	102	2	B71503	Ig light chain C r	344	6	0.8	138	2	H90775	host-nuclease inh
272	6	0.8	102	2	B34509	matrix Gla protein	345	6	0.8	138	2	B83366	hypothetical prote
273	6	0.8	103	1	GERBM1	PTS system, cellob	346	6	0.8	139	1	RKQMSX	ribulose-bisphosph
274	6	0.8	103	2	AE4140	unknown protein en	347	6	0.8	139	2	S78312	Ig light chain C r
275	6	0.8	103	2	AE5821	hypothetical prote	348	6	0.8	140	2	A34509	hypothetical prote
276	6	0.8	103	2	EP0973	hypothetical prote	349	6	0.8	140	2	S74052	hypothetical prote
277	6	0.8	103	2	AD0713	histone H4 - Oxytr	350	6	0.8	141	2	EE9260	hypothetical prote
278	6	0.8	104	2	J50154	histone H4 - Oxytr	351	6	0.8	141	2	DB7402	hypothetical prote
279	6	0.8	104	2	S14184	histone H4 (clone	352	6	0.8	141	2	S08523	hypothetical prote
280	6	0.8	104	2	S55062	protein L-asparyl	353	6	0.8	141	2	S34617	hypothetical prote
281	6	0.8	104	2	E97300	hypothetical prote	354	6	0.8	143	1	FEVTA	hypothetical prote
282	6	0.8	105	2	S62082	M-like protein em	355	6	0.8	144	1	JC4266	ferredoxin [Fe-4S
283	6	0.8	105	2	T23564	hypothetical prote	356	6	0.8	144	2	S23658	interleukin-3 prec
284	6	0.8	105	2	G69790	conserved hypotet	357	6	0.8	144	2	S32235	superoxide dismuta
285	6	0.8	109	2	G75484	hypothetical prote	358	6	0.8	145	2	T51740	ribosomal protein
286	6	0.8	109	2	D81358	probable integral	359	6	0.8	145	2	S14185	RNA helicase RH6 l
287	6	0.8	110	2	S70789	cagc protein precu	360	6	0.8	145	2	T10266	histone H4 (clone
288	6	0.8	110	2	E90806	probable curli pro	361	6	0.8	145	2	H70946	ribosomal protein
289	6	0.8	110	2	A85666	conserved hypotet	362	6	0.8	146	1	I46407	hypothetical prote
290	6	0.8	111	2	C69168	conserved hypotet	363	6	0.8	146	2	A86490	interleukin-3 prec
291	6	0.8	111	2	G72745	hypothetical prote	364	6	0.8	146	2	G71181	protein T32520.11
292	6	0.8	113	2	T44302	hypothetical prote	365	6	0.8	147	2	GB3301	hypothetical prote
293	6	0.8	113	2	D89762	hypothetical prote	366	6	0.8	148	2	C54128	conserved hypotet
294	6	0.8	115	2	S44112	Ig heavy chain V r	367	6	0.8	148	1	R5MTLA	ribosomal protein
295	6	0.8	116	2	S60818	M protein precuro	368	6	0.8	148	2	T51602	CAP binding protei
296	6	0.8	117	2	T12923	hypothetical prote	369	6	0.8	148	2	S25593	60S ribosomal prot
297	6	0.8	117	2	H84408	hypothetical prote	370	6	0.8	148	2	S60001	60S ribosomal prot
298	6	0.8	118	2	G84322	hypothetical prote	371	6	0.8	148	2	C54128	M-like protein em
299	6	0.8	119	2	E81272	hypothetical prote	372	6	0.8	148	2	T33937	hypothetical prote
300	6	0.8	120	2	T18059	hypothetical prote	373	6	0.8	149	1	R6NC7A	ribosomal protein
301	6	0.8	121	2	S61635	ribosomal protein	374	6	0.8	149	2	UC2392	ribosomal protein
302	6	0.8	121	2	D83624	probable two-compo	375	6	0.8	149	2	UC6202	ribosomal protein
303	6	0.8	122	1	S58649	ribosomal protein	376	6	0.8	149	2	D84292	transcription regu
304	6	0.8	122	2	AC1848	hypothetical prote	377	6	0.8	149	2	F82269	conserved hypotet
305	6	0.8	123	2	AG3530	hypothetical prote	378	6	0.8	149	2	B72694	hypothetical prote
306	6	0.8	126	2	G95154	conserved hypotet	379	6	0.8	151	1	MBBPA7	N-acetyl[muramoyl-L
307	6	0.8	127	2	F81021	hypothetical prote	380	6	0.8	151	2	S07306	N-acetyl[muramoyl-L
308	6	0.8	129	2	G69313	SSU ribosomal prot	381	6	0.8	151	2	A96493	hypothetical prote
309	6	0.8	129	2	S69854	hypothetical prote	382	6	0.8	151	2	T32957	hypothetical prote
310	6	0.8	129	2	T44125	hypothetical prote	383	6	0.8	151	2	S23215	hypothetical prote
311	6	0.8	129	2	G95303	hypothetical prote	384	6	0.8	151	2	BE9077	hypothetical prote
312	6	0.8	129	2	AC3607	hypothetical prote	385	6	0.8	152	2	T44544	conserved hypotet
313	6	0.8	130	1	R3LV11	ribosomal protein	386	6	0.8	153	2	C98238	hypothetical. 16.4k
314	6	0.8	130	2	S17719	H+-transporting tw	387	6	0.8	153	2	AH3047	conserved hypotet
315	6	0.8	130	2	S73215	ribosomal protein	388	6	0.8	153	2	H85638	hypothetical prote
316	6	0.8	130	2	S78255	ribosomal protein	389	6	0.8	154	2	T17124	hypothetical prote
317	6	0.8	130	2	T06945	ribosomal protein	390	6	0.8	154	2	D85093	hypothetical prote
318	6	0.8	130	2	T07518	ribosomal protein	391	6	0.8	154	2	A70508	hypothetical prote
319	6	0.8	130	2	C71023	hypothetical prote	392	6	0.8	154	2	C91171	probable beta-hydr
320	6	0.8	130	2	T43606	yope chaparon b/c	393	6	0.8	154	2	C86017	probable beta-hydr
321	6	0.8	130	2	T31161	DNA primase homolo	394	6	0.8	155	2	T03463	hypothetical prote

395 6 0.8 156 1 R3TW7 ribosomal protein 468
 396 6 0.8 156 1 JG0008 ribosomal protein 469
 397 6 0.8 156 2 F69699 ribosomal protein 470
 398 6 0.8 156 2 T44379 ribosomal protein 471
 399 6 0.8 156 2 B81234 30S ribosomal prot 472
 400 6 0.8 156 2 AF1406 ribosomal protein 473
 401 6 0.8 156 2 AF1782 ribosomal protein 474
 402 6 0.8 156 2 AF2210 hypothetical prote 475
 403 6 0.8 157 2 S23850 ribosomal protein 476
 404 6 0.8 157 2 B87117 conserved hypothet 477
 405 6 0.8 157 2 T24917 hypothetical prote 478
 406 6 0.8 158 2 G82494 conserved hypothet 479
 407 6 0.8 158 2 C32055 nifX protein - Syn 480
 408 6 0.8 159 2 S75802 kdtB protein - Azo 481
 409 6 0.8 159 2 T48837 hypothetical prote 482
 410 6 0.8 159 2 B83339 hypothetical prote 483
 411 6 0.8 160 2 A71723 ribosomal protein 484
 412 6 0.8 160 2 E97721 30S ribosomal prot 485
 413 6 0.8 160 2 E95882 hypothetical trans 486
 414 6 0.8 160 2 T16043 hypothetical prote 487
 415 6 0.8 161 2 T01651 hypoxically induce 488
 416 6 0.8 162 2 T11802 NADH2 dehydrogenas 489
 417 6 0.8 162 2 T45343 hypothetical prote 490
 418 6 0.8 162 2 C72665 hypothetical prote 491
 419 6 0.8 163 2 F81374 hypothetical prote 492
 420 6 0.8 163 2 T06903 ubiquinol-cytochro 493
 421 6 0.8 164 2 JCS572 hypothetical 18K p 494
 422 6 0.8 164 2 S35021 nolB protein - Rhi 495
 423 6 0.8 165 2 T21313 hypothetical prote 496
 424 6 0.8 165 2 F97519 hypothetical prote 497
 425 6 0.8 165 2 AH2738 acetyltransferase 498
 426 6 0.8 166 2 E71342 probable deoxycyti 499
 427 6 0.8 166 2 C90029 hypothetical prote 500
 428 6 0.8 167 1 JC1102 endothelin 3 precu 501
 429 6 0.8 167 2 B53293 flm3 region hypoth 502
 430 6 0.8 167 2 S70824 hypothetical prote 503
 431 6 0.8 167 2 F97238 probable acetyltra 504
 432 6 0.8 168 1 CYFGG2 gamma-crystallin I 505
 433 6 0.8 168 2 C75425 hypothetical prote 506
 434 6 0.8 169 2 T01783 gamma-2-crystallin 507
 435 6 0.8 169 2 AB1010 formylmethionine d 508
 436 6 0.8 169 2 C75427 hypothetical prote 509
 437 6 0.8 170 2 S77192 general secretion 510
 438 6 0.8 170 2 F83325 hypothetical prote 511
 439 6 0.8 170 2 C81967 probable acetyltra 512
 440 6 0.8 171 2 E87417 transcription regu 513
 441 6 0.8 172 2 F97320 acetyltransferase 514
 442 6 0.8 173 2 G86905 hypothetical prote 515
 443 6 0.8 173 2 H82671 hypothetical prote 516
 444 6 0.8 174 2 AF1668 hypothetical prote 517
 445 6 0.8 174 2 AH1296 hypothetical prote 518
 446 6 0.8 175 1 VGXRNH glycoprotein NCPV5 519
 447 6 0.8 175 1 VGXR66 glycoprotein NCPV5 520
 448 6 0.8 175 1 VGXRBR glycoprotein NCPV5 521
 449 6 0.8 175 1 VGXR65 glycoprotein NCPV5 522
 450 6 0.8 175 2 T49053 glycoprotein NCPV5 523
 451 6 0.8 175 2 C97138 hypothetical prote 524
 452 6 0.8 176 2 C87070 conserved hypothet 525
 453 6 0.8 176 2 H70577 hypothetical prote 526
 454 6 0.8 177 2 T31242 hypothetical prote 527
 455 6 0.8 177 2 S54778 NR-13 protein - qu 528
 456 6 0.8 177 2 F95912 hypothetical prote 529
 457 6 0.8 178 2 F82952 ATP synthase delta 530
 458 6 0.8 178 2 S24298 chorion protein - 531
 459 6 0.8 178 2 AB0870 outer membrane pro 532
 460 6 0.8 179 2 T20206 hypothetical prote 533
 461 6 0.8 180 2 T02388 hypothetical prote 534
 462 6 0.8 181 1 TXSPM thioredoxin m prec 535
 463 6 0.8 181 2 T24914 hypothetical prote 536
 464 6 0.8 181 2 AE0714 probable chorismat 537
 465 6 0.8 181 2 F82750 outer membrane pro 538
 466 6 0.8 182 2 B53311 acetyl-CoA carboxy 539
 467 6 0.8 182 2 A12437 biotin carboxyl ca 540

182 2 S03259 methyl coenzyme M 2
 183 2 C75386 ribosome recycling 2
 184 2 G64786 ybcL protein - Esc 2
 185 2 T35733 hypothetical prote 2
 186 2 T13186 hypothetical prote 2
 187 2 JQ1623 envelope-associate 1
 188 2 JQ1987 hypothetical 21.0K 1
 189 2 B83727 hypothetical prote 2
 190 2 S51022 M21 chain - chicke 2
 191 2 E97353 hypothetical prote 2
 192 2 T00427 ribosomal protein 2
 193 2 H84916 60S ribosomal prot 2
 194 2 E90191 conserved hypothet 2
 195 2 A71328 probable survival 2
 196 1 R5RT81 ribosomal protein 1
 197 1 JN0779 ribosomal protein 1
 198 1 R5XL6A ribosomal protein 1
 199 1 R5XL14 ribosomal protein 1
 200 1 S38352 ribosomal protein 1
 201 1 IVMSA5 interferon alpha-1 1
 202 1 H96903 transcription regu 1
 203 1 H96903 interferon alpha-2 1
 204 1 JHMSA2 interferon alpha-1 1
 205 1 JH0468 interferon alpha-1 1
 206 2 A24401 interferon alpha-1 2
 207 2 F72206 hypothetical prote 2
 208 2 F72731 hypothetical prote 2
 209 2 AD2530 hypothetical prote 2
 210 2 G70102 Holliday junction 2
 211 2 AF2356 hypothetical prote 2
 212 2 S37970 succinate dehydrog 2
 213 2 G82265 syd protein VC0903 2
 214 2 A72760 hypothetical prote 2
 215 1 S47726 outer-membrane lip 1
 216 1 S35441 serine-tRNA ligase 1
 217 2 A91176 outer membrane pro 2
 218 2 E86022 outer membrane pro 2
 219 2 T05293 hypothetical prote 2
 220 2 T23192 hypothetical prote 2
 221 2 T46404 hypothetical prote 2
 222 2 A82160 hypothetical prote 2
 223 2 B87342 hypothetical prote 2
 224 2 I49153 cardiotrophin-1 - 2
 225 2 JC4845 conserved hypothet 2
 226 2 T08332 mitochondrial ribo 2
 227 2 T48718 60S ribosomal prot 2
 228 2 B90112 hypothetical prote 2
 229 2 T19699 hypothetical prote 2
 230 2 I40179 heme acquisition p 2
 231 2 G83218 phosphoglycerate m 2
 232 2 F95113 hypothetical prote 2
 233 2 F84087 ribosomal protein 2
 234 2 E95024 50S ribosomal prot 2
 235 2 E97895 transcription regu 2
 236 2 F84203 hypothetical prote 2
 237 2 F84518 uracil phosphoribo 2
 238 2 AE0344 uracil phosphoribo 2
 239 2 AH0818 l-lactate permease 2
 240 2 AB3406 probable orotate p 2
 241 2 S44298 conserved hypothet 2
 242 2 H81651 conserved hypothet 2
 243 2 T19847 conserved hypothet 2
 244 2 D83034 conserved hypothet 2
 245 2 JQ1554 mitochondrial inne 2
 246 2 AE0187 probable aldehyde 2
 247 2 J71876 oxygen-insensitive 2
 248 2 B64639 vard protein - pse 2
 249 2 I40540 conserved hypothet 2
 250 2 F64609 hypothetical prote 2
 251 2 S28673 probable thymidyla 2
 252 2 E87005 hypothetical prote 2
 253 2 T49785 L-isoeaspartate pro 2
 254 2 F83193 conserved hypothet 2
 255 2 H82803 transcription term 2
 256 2 AB2026

541	6	0.8	213	2	S01766	GTP-binding protei	614	6	0.8	238	1	UC2297	CD3 antigen - rab
542	6	0.8	213	2	E34323	GTP-binding protei	615	6	0.8	238	1	S43511	CD63/ME491 antigen
543	6	0.8	213	2	G75521	ABC transporter, A	616	6	0.8	238	1	A46508	CD63/ME491 antigen
544	6	0.8	213	2	S43723	lactase (EC 3.2.1.	617	6	0.8	238	2	T10771	NADPH-ferrithemopo
545	6	0.8	213	2	AF0002	molycoprotein-guan	618	6	0.8	238	2	S70468	agglutinin (WBA 1)
546	6	0.8	214	2	T05176	hypothetical prote	619	6	0.8	238	2	D83847	two-component resp
547	6	0.8	214	2	S27523	hypothetical prote	620	6	0.8	238	2	A72374	hypothetical prote
548	6	0.8	215	2	T22572	hypothetical prote	621	6	0.8	238	2	H95105	pilin gene inverti
549	6	0.8	215	2	S75345	hypothetical prote	622	6	0.8	239	2	F84189	hypothetical prote
550	6	0.8	215	2	AD2407	heterocyst diffe	623	6	0.8	239	2	D96587	hypothetical prote
551	6	0.8	216	2	H87547	proteocathecate 3,	624	6	0.8	239	2	T15222	hypothetical prote
552	6	0.8	217	1	A65026	uracil phosphoribo	625	6	0.8	239	2	A72623	hypothetical prote
553	6	0.8	217	1	W0H02C	homeotic protein H	626	6	0.8	240	2	A41797	Ig light chain - s
554	6	0.8	217	2	H91048	uracil phosphoribo	627	6	0.8	240	2	A10941	probable GntR-fam
555	6	0.8	217	2	D85893	uracil phosphoribo	628	6	0.8	241	2	AC2646	conserved hypotet
556	6	0.8	217	2	T31875	hypothetical prote	629	6	0.8	241	2	AC1918	hypothetical prote
557	6	0.8	218	2	F86844	hypothetical prote	630	6	0.8	241	2	S74620	hypothetical prote
558	6	0.8	218	2	B97428	hypothetical prote	631	6	0.8	241	2	A42299	hypothetical prote
559	6	0.8	218	2	A34445	25K calcium-bindin	632	6	0.8	242	2	DB1161	conserved hypotet
560	6	0.8	218	2	G70654	hypothetical prote	633	6	0.8	242	2	H71806	probable outer mem
561	6	0.8	219	2	A44503	p25 protein - beet	634	6	0.8	244	2	A55541	heterocyst maturat
562	6	0.8	219	2	G64009	hypothetical prote	635	6	0.8	244	2	A12269	heterocyst specifi
563	6	0.8	220	2	T44702	hypothetical prote	636	6	0.8	244	2	A69180	precocorin-3-methyl
564	6	0.8	220	2	S75751	hypothetical prote	637	6	0.8	244	2	A41021	NAD+-asparagine AD
565	6	0.8	221	2	S20964	ribosomal protein	638	6	0.8	245	1	C65210	hypothetical 26..3
566	6	0.8	221	2	G84169	hypothetical prote	639	6	0.8	245	1	A72218	guFA protein - The
567	6	0.8	221	2	S54452	hypothetical prote	640	6	0.8	245	2	C91255	hypothetical prote
568	6	0.8	222	2	S26994	cellulose 1,4-beta	641	6	0.8	245	2	G86095	hypothetical prote
569	6	0.8	222	2	S20029	homeotic protein H	642	6	0.8	245	2	B95943	probable choline u
570	6	0.8	222	2	A81944	hypothetical prote	643	6	0.8	245	2	168673	gene X123 protein
571	6	0.8	222	2	I39192	gene HOXA1 protein	644	6	0.8	246	2	C70316	hypothetical prote
572	6	0.8	222	2	G84175	regulatory protein	645	6	0.8	246	2	S27727	hypothetical prote
573	6	0.8	223	1	A41965	transcription regu	646	6	0.8	246	2	A10920	probable UDP-N-ace
574	6	0.8	223	2	A81285	probable triose-ph	647	6	0.8	246	2	AH3072	transcription regu
575	6	0.8	223	2	B90829	transcription regu	648	6	0.8	248	1	J01682	infected cell prot
576	6	0.8	223	2	H85686	transcription regu	649	6	0.8	248	2	AC1657	glucose 1-dehydrog
577	6	0.8	223	2	AE0199	response regulator	650	6	0.8	248	2	AH1285	glucose 1-dehydrog
578	6	0.8	224	1	RGBBFT	transcription regu	651	6	0.8	248	2	G81274	probable cyclase C
579	6	0.8	224	2	AH0646	transcription regu	652	6	0.8	248	2	F72087	conserved hypotet
580	6	0.8	224	2	T10660	phosphatase II pro	653	6	0.8	248	2	D66537	CT056 hypotet
581	6	0.8	224	2	T32606	hypothetical prote	654	6	0.8	248	2	B64703	outer membrane pro
582	6	0.8	225	2	G83371	probable amino aci	655	6	0.8	248	2	E71816	probable outer mem
583	6	0.8	225	2	T02660	germin-like protei	656	6	0.8	249	2	A82796	hypothetical prote
584	6	0.8	225	2	T04210	hypothetical prote	657	6	0.8	249	2	T05424	hypothetical prote
585	6	0.8	225	2	F89795	hypothetical prote	658	6	0.8	250	2	H82079	protein disulfide-
586	6	0.8	226	2	S70467	agglutinin (WBA 1)	659	6	0.8	250	2	T10663	hypothetical prote
587	6	0.8	226	2	JC5327	adhesin complex 25	660	6	0.8	250	2	T35927	conserved hypotet
588	6	0.8	227	2	S01260	M protein precurs	661	6	0.8	250	2	A46957	NAD+-asparagine AD
589	6	0.8	228	2	S67199	hypothetical prote	662	6	0.8	251	2	B64102	ribosomal protein
590	6	0.8	229	2	H75205	hypothetical prote	663	6	0.8	251	2	G81279	enterocheilin uptak
591	6	0.8	230	2	T06709	protein-L-isoaspar	664	6	0.8	251	2	F72114	conserved hypotet
592	6	0.8	231	2	S37108	cuticlin 2 - Caen	665	6	0.8	251	2	A86508	ACR family [import
593	6	0.8	231	2	AE1907	hypothetical prote	666	6	0.8	251	2	T32308	hypothetical prote
594	6	0.8	231	2	T44086	transcription regu	667	6	0.8	251	2	B64048	hypothetical prote
595	6	0.8	231	2	T02765	glutathione trans	668	6	0.8	252	2	S69786	pikR protein - Sal
596	6	0.8	231	2	C86665	amino acid ABC tra	669	6	0.8	252	2	AC0849	pathogenicity 1 is
597	6	0.8	231	2	B72100	hypothetical prote	670	6	0.8	252	2	D66462	ATG1-like protein,
598	6	0.8	231	2	B66524	hypothetical prote	671	6	0.8	252	2	B84205	hypothetical prote
599	6	0.8	232	1	A54361	venombin A (EC 3.4	672	6	0.8	253	2	T41139	proteasome compone
600	6	0.8	232	2	A87083	conserved hypotet	673	6	0.8	253	2	169732	stationary-phase s
601	6	0.8	233	2	I51383	Ig lambda chain -	674	6	0.8	253	2	F91078	survival protein f
602	6	0.8	233	2	A10109	probable response	675	6	0.8	253	2	G85923	stationary-phase s
603	6	0.8	233	2	T35251	probable endonuc	676	6	0.8	253	2	AB0856	stationary-phase s
604	6	0.8	234	2	C70156	hypothetical prote	677	6	0.8	253	2	H70780	hypothetical prote
605	6	0.8	236	2	T27514	hypothetical prote	678	6	0.8	253	2	AC2627	hypothetical prote
606	6	0.8	236	2	B86281	7alpa-cephem-meth	679	6	0.8	254	2	JC5004	hypothetical prote
607	6	0.8	236	2	C95387	hypothetical prote	680	6	0.8	254	2	AH0407	attacin A precurs
608	6	0.8	236	2	D72556	hypothetical prote	681	6	0.8	254	2	T35366	probable membrane
609	6	0.8	237	2	B69838	transcription regu	682	6	0.8	255	1	G69495	probable membrane
610	6	0.8	237	2	C95199	nitroductase fam	683	6	0.8	255	1	A28169	venombin A (EC 3.4
611	6	0.8	237	2	A98066	NADPH-flavin oxido	684	6	0.8	255	2	S64574	probable membrane
612	6	0.8	237	2	B64643	hypothetical prote	685	6	0.8	255	2	G71007	thiamin biosynthes
613	6	0.8	238	1	I38016	melanoma-associate	686	6	0.8	255	2	A82671	conserved hypotet

687 6 0.8 256 2 F97555 triosephosphate is
688 6 0.8 256 2 AH2775 triosephosphate is
689 6 0.8 256 2 T12999 aquaporin homolog
690 6 0.8 256 2 S11877 chlorophyll a/b-bi
691 6 0.8 257 2 G95252 PTS system, IIC co
692 6 0.8 257 2 E98117 hypothetical prote
693 6 0.8 257 2 H84713 hypothetical prote
694 6 0.8 257 2 F75476 probable uroporphyr
695 6 0.8 257 2 T23332 hypothetical prote
696 6 0.8 257 2 D70341 hypothetical prote
697 6 0.8 258 2 S61656 hypothetical prote
698 6 0.8 258 2 S73803 MG256 homolog H91_
699 6 0.8 258 2 B97540 transmembrane regu
700 6 0.8 258 2 AD2759 transcription regu
701 6 0.8 258 2 H72202 ROK family protein
702 6 0.8 259 2 T34637 probable type IV p
703 6 0.8 259 2 S41890 hypothetical prote
704 6 0.8 260 2 S22850 ERs1 protein - yea
705 6 0.8 261 2 C87204 chitoesterase limpo
706 6 0.8 261 2 AC3646 hydroxypyruvate is
707 6 0.8 261 2 T24532 hypothetical prote
708 6 0.8 261 2 E75626 hypothetical prote
709 6 0.8 261 2 S44776 C30A5 protein - Ca
710 6 0.8 263 2 D71281 probable pyrrolina
711 6 0.8 263 2 F92311 survival protein S
712 6 0.8 263 2 S64472 hypothetical prote
713 6 0.8 264 2 A84957 methionyl aminopep
714 6 0.8 264 2 AD3353 probable membrane
715 6 0.8 264 2 F70969 hypothetical prote
716 6 0.8 264 2 D97646 probable ABC-trans
717 6 0.8 264 2 AD2870 hypothetical prote
718 6 0.8 265 2 AF2007 hypothetical prote
719 6 0.8 266 1 ENSAC1 enterotoxin C-1 pr
720 6 0.8 266 2 E69083 thiamin biosynthes
721 6 0.8 266 2 C95288 probable ABC trans
722 6 0.8 267 1 A46493 metastasis suppres
723 6 0.8 267 2 B72408 conserved hypoteth
724 6 0.8 267 2 T15595 hypothetical prote
725 6 0.8 267 2 T72415 zinc ABC transport
726 6 0.8 267 2 H97299 proB [imported] -
727 6 0.8 268 2 AG2594 enoyl-(acyl)-carrie
728 6 0.8 268 2 S68826 pancreatic elastas
729 6 0.8 268 2 S68825 pancreatic elastas
730 6 0.8 268 2 S67583 KNH1 protein - yea
731 6 0.8 268 2 S72968 cytotoxin/hemolysi
732 6 0.8 268 2 E70502 probable tlyA prot
733 6 0.8 268 2 B75279 survival protein S
734 6 0.8 268 2 T35568 probable lipoprote
735 6 0.8 268 2 B98223 branched-chain ami
736 6 0.8 268 2 AF3063 hypothetical prote
737 6 0.8 269 2 B44499 major intrinsic pr
738 6 0.8 269 2 A41616 erythrocyte integr
739 6 0.8 269 2 JG1320 water channel prot
740 6 0.8 269 2 I52366 uterine water chan
741 6 0.8 269 2 H87078 cytotoxin/hemolysi
742 6 0.8 269 2 C91030 probable structura
743 6 0.8 269 2 D85874 probable structura
744 6 0.8 269 2 A00804 probable membrane
745 6 0.8 269 2 E98165 hypothetical prote
746 6 0.8 269 2 AB3122 hypothetical prote
747 6 0.8 269 2 S08346 probable structura
748 6 0.8 270 2 H97376 enoyl-(acyl)-carrie
749 6 0.8 270 2 AF1172 mannose-specific p
750 6 0.8 270 2 F72381 D-tagatose 3-epime
751 6 0.8 271 2 AG1529 mannose-specific p
752 6 0.8 271 2 B84861 50S ribosomal prot
753 6 0.8 271 2 JC2348 water channel prot
754 6 0.8 271 2 S75967 hypothetical prote
755 6 0.8 271 2 T21777 hypothetical prote
756 6 0.8 271 2 AF2234 glutamine-binding
757 6 0.8 272 2 AH3496 enoyl-[acyl]-carrie
758 6 0.8 272 2 C72498 probable stress pr
759 6 0.8 272 2 B75293 amino acid ABC tra

760 6 0.8 272 2 E70862 hypothetical prote
761 6 0.8 272 2 H87395 hypothetical prote
762 6 0.8 272 2 C96742 unknown protein F1
763 6 0.8 272 2 H71124 hypothetical prote
764 6 0.8 272 2 A72512 probable hypoxanth
765 6 0.8 272 2 G89856 conserved hypoteth
766 6 0.8 273 2 T08645 yfjQ protein - Esc
767 6 0.8 273 2 T40611 hypothetical prote
768 6 0.8 273 2 S21731 signal recognition
769 6 0.8 274 2 D95165 NH(3)-dependent NA
770 6 0.8 274 2 C98031 NAD synthase (glut
771 6 0.8 274 2 TQ0686 nosiheptide resist
772 6 0.8 274 2 T20435 hypothetical prote
773 6 0.8 274 2 E83909 oligopeptide ABC t
774 6 0.8 275 2 T17966 ribonuclease III h
775 6 0.8 275 2 F70191 rare lipoprotein A
776 6 0.8 275 2 S76233 transcription term
777 6 0.8 276 2 D41044 octopine-binding p
778 6 0.8 276 2 G75358 hypothetical prote
779 6 0.8 276 2 T29689 hypothetical prote
780 6 0.8 276 2 D90595 hypothetical prote
781 6 0.8 277 2 B83353 probable binding-p
782 6 0.8 277 2 B87651 conserved hypoteth
783 6 0.8 278 2 D64750 yafZ protein - Esc
784 6 0.8 278 2 G75177 dipeptide transpor
785 6 0.8 278 2 S09662 hypothetical prote
786 6 0.8 278 2 A13633 hypothetical prote
787 6 0.8 279 2 C70458 diaminopimelate ep
788 6 0.8 279 2 AG3603 maltose transport
789 6 0.8 279 2 C87444 conserved hypoteth
790 6 0.8 279 2 E75573 transcription regu
791 6 0.8 280 2 G72428 oligopeptide ABC t
792 6 0.8 280 2 S64246 hypothetical prote
793 6 0.8 280 2 S48903 SOL3 protein - yea
794 6 0.8 280 2 T30080 hypothetical prote
795 6 0.8 281 2 AC3471 cysQ protein limpo
796 6 0.8 281 2 T22280 hypothetical prote
797 6 0.8 281 2 T39199 C2H2-150 - human
798 6 0.8 282 2 JS0168 collagen col-8 - C
799 6 0.8 282 2 T16036 cuticle collagen c
800 6 0.8 283 2 A64174 hypothetical prote
801 6 0.8 284 2 T52062 PRAG-like protein
802 6 0.8 285 2 E75317 glycerophosphoryl
803 6 0.8 285 2 A97561 hypothetical prote
804 6 0.8 286 2 H64054 site-specific DNA-
805 6 0.8 286 2 H88175 protein T24H7.1 li
806 6 0.8 286 2 T34096 hypothetical prote
807 6 0.8 286 2 B95011 hypothetical prote
808 6 0.8 287 2 T22605 hypothetical prote
809 6 0.8 288 2 H84757 hypothetical prote
810 6 0.8 289 2 H95929 probable hydroxyme
811 6 0.8 289 2 T02839 probable membrane
812 6 0.8 289 2 A99367 hypothetical prote
813 6 0.8 289 2 E81259 hypothetical prote
814 6 0.8 290 2 JX0175 lectin precursor -
815 6 0.8 290 2 T09630 transcription regu
816 6 0.8 290 2 S53904 hypothetical prote
817 6 0.8 290 2 C90382 membrane conserved
818 6 0.8 290 2 H64431 glycosyl transfera
819 6 0.8 291 2 T04122 preprotein translo
820 6 0.8 291 2 T11983 thiamin biosynthes
821 6 0.8 291 2 E95106 carbon-nitrogen hy
822 6 0.8 291 2 G97974 beta-ureidopropion
823 6 0.8 291 2 B90256 conserved hypoteth
824 6 0.8 291 2 A72341 hypothetical prote
825 6 0.8 291 2 G95139 geranyltransferrin
826 6 0.8 291 2 E84646 geranyltransferrin
827 6 0.8 291 2 F98007 geranyltransferrin
828 6 0.8 292 1 B35124 biphényl-2,3-diol
829 6 0.8 292 2 G75399 N-acetylglutamate
830 6 0.8 292 2 T09531 2-dehydro-3-deoxy-
831 6 0.8 293 2 AC2215 UTP-glucose-1-phos
832 6 0.8 293 2 A89818 hypothetical prote

833	6	0.8	294	2	B87210	hypothetical prote	906	6	0.8	311	2	B90715	probable tRNA synt
834	6	0.8	294	2	AB0115	probable carbon-ni	907	6	0.8	311	2	B85565	probable tRNA synt
835	6	0.8	294	2	B86870	hypothetical prote	908	6	0.8	311	2	AB0583	probable nucleosid
836	6	0.8	294	2	F64337	hypothetical prote	909	6	0.8	312	2	C70475	cytochrome c bioge
837	6	0.8	294	2	B69759	hypothetical prote	910	6	0.8	312	2	G71146	hypothetical prote
838	6	0.8	294	2	D87505	geranyltranscranf	911	6	0.8	312	2	G84973	tRNA pseudouridine
839	6	0.8	294	2	G95861	probable inositol	912	6	0.8	312	2	A42791	conserved hypotnet
840	6	0.8	295	2	AB2156	hypothetical prote	913	6	0.8	312	2	T20932	hypothetical prote
841	6	0.8	296	2	S72943	diaminopimelate ep	914	6	0.8	313	2	B69588	L-arabinose transp
842	6	0.8	296	2	T48926	bzip transcription	915	6	0.8	313	2	B64896	probable membrane
843	6	0.8	296	2	T38993	microtubule-associ	916	6	0.8	313	2	A85734	probable transport
844	6	0.8	296	2	T27768	hypothetical prote	917	6	0.8	313	2	P90864	probable transport
845	6	0.8	296	2	H81726	UDP-N-acetylenolpy	918	6	0.8	313	2	S60713	polygalacturonase-
846	6	0.8	297	1	ROBO	thiosulfate sulfur	919	6	0.8	313	2	S30954	minor tail protein
847	6	0.8	297	1	JC5286	geranyltrancranf	920	6	0.8	313	2	D75403	hypothetical prote
848	6	0.8	297	1	JX0257	biphenyl-2,3-diol	921	6	0.8	313	2	A48903	beta-lactamase - P
849	6	0.8	298	1	UN0815	biphenyl-2,3-diol	922	6	0.8	314	2	G95403	probable ABC trans
850	6	0.8	298	1	T20841	hypothetical prote	923	6	0.8	314	2	H70840	probable regulator
851	6	0.8	298	2	T20841	hypothetical prote	924	6	0.8	314	2	H70801	hypothetical prote
852	6	0.8	299	2	T01806	hypothetical prote	925	6	0.8	314	2	H89849	conserved hypotnet
853	6	0.8	299	2	C90057	hypothetical prote	926	6	0.8	314	2	AE3193	conserved hypotnet
854	6	0.8	300	2	A71645	protein p34 (p34)	927	6	0.8	314	2	G83487	hypothetical prote
855	6	0.8	300	2	S72756	cytochrome O ubiqn	928	6	0.8	314	2	A70503	hypothetical prote
856	6	0.8	300	2	B69053	hypothetical prote	929	6	0.8	315	2	H97570	hypothetical 34.4k
857	6	0.8	300	2	T00274	hypothetical prote	930	6	0.8	315	2	I50706	transcription fact
858	6	0.8	300	2	H86358	zinc finger protei	931	6	0.8	316	2	S50336	NADH2 dehydrogenas
859	6	0.8	301	1	RDHUB5	cytochrome-b5 redu	932	6	0.8	316	2	B90282	hypothetical prote
860	6	0.8	301	1	H95308	probable ABC trans	933	6	0.8	316	2	D83406	probable transmemb
861	6	0.8	301	2	S72642	probable ABC-type	934	6	0.8	316	2	T08694	hypothetical prote
862	6	0.8	301	2	A88931	protein R11G11.15	935	6	0.8	316	2	T28072	hypothetical prote
863	6	0.8	301	2	A98214	hypothetical prote	936	6	0.8	316	2	G82696	rod shape-determin
864	6	0.8	302	2	T34398	hypothetical prote	937	6	0.8	317	2	F71139	hypothetical prote
865	6	0.8	302	2	T09066	probable palmitoyl	938	6	0.8	317	2	E75421	conserved hypotnet
866	6	0.8	303	1	B36227	urate oxidase (EC	939	6	0.8	317	2	G83033	vanillate O-demeth
867	6	0.8	303	2	A45640	phosphoprotein pho	940	6	0.8	317	2	F89349	D-ribose-binding p
868	6	0.8	303	2	S74949	hypothetical prote	941	6	0.8	317	2	A12932	hypothetical prote
869	6	0.8	303	2	B84296	hypothetical prote	942	6	0.8	317	2	AE1932	hypothetical prote
870	6	0.8	303	2	A71819	hypothetical prote	943	6	0.8	318	2	C82231	chloroedoxin reduct
871	6	0.8	303	2	F64701	hypothetical prote	944	6	0.8	318	2	B72676	hypothetical prote
872	6	0.8	303	2	H84758	hypothetical prote	945	6	0.8	318	2	A99510	hypothetical prote
873	6	0.8	303	2	F84401	hypothetical prote	946	6	0.8	318	2	AD3200	conserved hypotnet
874	6	0.8	303	2	S74517	hypothetical prote	947	6	0.8	319	2	S61421	aldose reductase h
875	6	0.8	303	2	C97463	hypothetical prote	948	6	0.8	319	2	C87447	aldose reductase-x
876	6	0.8	304	1	WZBEA9	33.2K tegument pro	949	6	0.8	320	1	S15024	aldose reductase-r
877	6	0.8	304	2	B82089	transcription regu	950	6	0.8	320	2	J02253	aldehyde reductase
878	6	0.8	304	2	AG1182	hypothetical prote	951	6	0.8	320	2	T48188	aldose reductase-1
879	6	0.8	304	2	AG0759	pdu/cob regulatory	952	6	0.8	320	2	B70947	probable moxR3 pro
880	6	0.8	305	2	F72574	probable cytochrom	953	6	0.8	320	2	AH0067	conserved hypotnet
881	6	0.8	305	2	H75061	hypothetical prote	954	6	0.8	320	2	F75348	hypothetical prote
882	6	0.8	306	2	S19997	hypothetical prote	955	6	0.8	320	2	S61586	probable membrane
883	6	0.8	306	2	T32801	hypothetical prote	956	6	0.8	321	2	H86981	probable prochrome
884	6	0.8	307	2	D87688	hypothetical prote	957	6	0.8	321	2	I18238	transcription fact
885	6	0.8	307	2	S51373	tau-protein kinase	958	6	0.8	322	2	UN0419	DNA-directed RNA p
886	6	0.8	307	2	S50850	cyclin-dependent k	959	6	0.8	322	2	F83140	thiamine-phosphate
887	6	0.8	307	2	S50861	cyclin-dependent k	960	6	0.8	322	2	T48460	MADS-box protein-1
888	6	0.8	307	2	C83188	probable ATP-bindi	961	6	0.8	323	2	T01626	peroxidase (EC 1.1
889	6	0.8	308	2	B82394	hypothetical prote	962	6	0.8	323	2	A72508	probable cobalamin
890	6	0.8	308	2	T08796	tropomyosin - huma	963	6	0.8	323	2	F84423	hypothetical prote
891	6	0.8	308	2	C83586	probable transcrip	964	6	0.8	323	2	H90119	hypothetical prote
892	6	0.8	308	2	T14739	hypothetical prote	965	6	0.8	324	1	B69553	methanol dehydroge
893	6	0.8	309	2	D75008	methanol dehydroge	966	6	0.8	324	2	C69261	2-hydroxyhepta-2,4
894	6	0.8	309	2	F72316	hypothetical prote	967	6	0.8	324	2	T28032	hypothetical prote
895	6	0.8	309	2	B38545	hypothetical prote	968	6	0.8	324	2	D95218	iron-compound ABC
896	6	0.8	309	2	F71824	hypothetical prote	969	6	0.8	324	2	B98082	hypothetical prote
897	6	0.8	310	2	A11059	carbamate kinase (970	6	0.8	324	2	G81330	probable phosphata
898	6	0.8	310	2	B87242	probable LysR faml	971	6	0.8	325	2	AE82073	GTP-binding protei
899	6	0.8	310	2	JC4343	uridine phosphoryl	972	6	0.8	325	2	A13635	thiamine biosynthe
900	6	0.8	310	2	H82439	hypothetical prote	973	6	0.8	325	2	AE3575	transcription regu
901	6	0.8	310	2	T29731	hypothetical prote	974	6	0.8	325	2	C91138	hypothetical prote
902	6	0.8	310	2	D71679	UDP-n-acetylenolpy	975	6	0.8	325	2	F85983	hypothetical prote
903	6	0.8	311	2	T50326	probable mitochond	976	6	0.8	325	2	A10904	probable membrane
904	6	0.8	311	2	D64240	methionyl-tRNA for	977	6	0.8	325	2	F55110	glycylprotein VP7 p
905	6	0.8	311	2	A64800	purine nucleosidas	978	6	0.8	326	1	J01442	

979 6 0.8 326 2 S14266 uracil-DNA glycosy
980 6 0.8 326 2 T43226 probable NS, N10-me
981 6 0.8 326 2 AG1316 protein-tyrosine/s
982 6 0.8 326 2 AG1688 protein-tyrosine/s
983 6 0.8 327 2 T10263 probable polygalac
984 6 0.8 327 2 B71146 probable glucose-1
985 6 0.8 327 2 D87645 sensor histidine k
986 6 0.8 328 2 A87410 myo-inositol 2-deh
987 6 0.8 329 2 T08262 hypothetical prote
988 6 0.8 329 2 S53494 RNA-binding protei
989 6 0.8 329 2 T00873 hypothetical prote
990 6 0.8 329 2 T20546 hypothetical prote
991 6 0.8 330 2 G75007 lps biosynthesis r
992 6 0.8 330 2 H64077 aspartate-ammonia
993 6 0.8 330 2 D95230 aspartate-ammonia
994 6 0.8 330 2 S40853 probable 2-keto-3-
995 6 0.8 330 2 C86080 2-keto-3-deoxy-D-g
996 6 0.8 330 2 C91233 2-keto-3-deoxy-D-g
997 6 0.8 330 2 S35439 transcription fact
998 6 0.8 330 2 AD1533 hypothetical prote
999 6 0.8 331 1 A70032 conserved hypothet
1000 6 0.8 331 2 F75074 hypothetical prote

ALIGNMENTS

RESULT 1
B81038
TonB-dependent receptor NMB1829 [imported] - Neisseria meningitidis (strain MC58 serog
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: B81038
R:Telletlin, H.; Saunders, N.J.; Heideberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Piza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: B81038
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <TET>
A:Cross-references: GB:AE002532; GB:AE002098; NID:g7227078; PIDN:AAF42164.1; PID:g722708
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1829

Query Match 23.8%; Score 172; DB 2; Length 708;
Best Local Similarity 100.0%; Pred. No. 6e-173;
Matches 172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTPGLRVLSNDDGRSSVYARGY 120
DB 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQARKTPGLRVLSNDDGRSSVYARGY 120
QY 121 EYSEYNIDGLPAQMOSINGTLPNLFADFVEVMRGPSGLFDSGEMGGIVNLVRKRPYKA 180
DB 121 EYSEYNIDGLPAQMOSINGTLPNLFADFVEVMRGPSGLFDSGEMGGIVNLVRKRPYKA 180
QY 181 FQGHAAAGFGTHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNR 232
DB 181 FQGHAAAGFGTHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNR 232
RESULT 2
AC2079
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2079

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2079
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073884.1; PID:g17131276; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2185
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 1.5%; Score 11; DB 2; Length 853;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 627 PKHSANLWTTY 637
DB 751 PKHSANLWTTY 761

RESULT 3
C83997
1-deoxyxylulose-5-phosphate synthase BH2779 [imported] - Bacillus halodurans (strain C-1
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C83997
R:Takami, R.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C83997
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-629 <STO>
A:Cross-references: GB:AP001516; GB:BA000004; NID:g10175192; PIDN:BA06498.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2779
C:Superfamily: hypothetical protein C2814

Query Match 1.2%; Score 9; DB 2; Length 629;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 GHYKIESGE 478
DB 299 GHYKIESGE 307

RESULT 4
D81976
probable ferric siderophore receptor protein NMA0575 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: D81976
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
Holroyd, S.; Jørgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: D81976
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-725 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83866.1; PID:g737931
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: NMA0575
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo

R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AC0547
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-696 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08819.1; PID:gl6501633; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY0396

Query Match 1.1% Score 8; DB 2; Length 696;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 PQSVSIIT 81
 |||||
 Db 54 PQSVSIIT 61

RESULT 10

A36942
 Fe(III)-pyochelin receptor fptA precursor - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 31-Dec-2000
 C:Accession: A36942; G83117
 R;Ankenbauer, R.G.; Quan, H.N.
 J. Bacteriol. 176, 307-319, 1994
 A:Title: FptA, the Fe(III)-pyochelin receptor of *Pseudomonas aeruginosa*: a phenolate siderophore
 A:Reference number: A36942; MUID:94117363; PMID:8288523
 A:Accession: A36942
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <ANK>
 A:Cross-references: GB:U03161; NID:9454352; PIDN:AAC43213.1; PID:9454353
 R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, M.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: G83117
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-720 <STO>
 A:Cross-references: GB:AE004839; GB:AE004091; NID:99950434; PIDN:AAG07609.1; GSPDB:GN00176
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: fptA; PA4221
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology <TNN>
 F:98-227/Domain: tonB-dependent receptor amino-terminal homology <TNN>
 F:446-720/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 1.1% Score 8; DB 2; Length 720;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 VNLVRKRP 177
 |||||
 Db 176 VNLVRKRP 183

RESULT 11

A80187
 probable iron-siderophore receptor YPO1537 [imported] - *Yersinia pestis* (strain CO92)
 C:Species: *Yersinia pestis*
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 C:Accession: A80187
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; Li, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: A80187
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-753 <KUR>
 A:Cross-references: GB:AL590842; PIDN:CAC90360.1; PID:gl5979580; GSPDB:GN00175
 C:Genetics:
 A:Gene: YPO1537
 C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology

Query Match 1.1% Score 8; DB 2; Length 753;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 NLVRKRP 178
 |||||
 Db 199 NLVRKRP 206

RESULT 12

T30917
 hypothetical protein D1007.15 - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T30917
 R;Davidson, S.; Rohlfing, T.
 submitted to the EMBL Data Library, August 1999
 A:Description: The sequence of C. elegans cosmid D1007.
 A:Reference number: Z20934
 A:Accession: T30917
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-775 <DAV>
 A:Cross-references: EMBL:AF003151; PIDN:AAB54226.2
 A:Experimental source: strain Bristol N2; clone D1007
 C:Genetics:
 A:Map position: I
 A:Introns: 44/3; 146/3; 224/1; 343/2; 422/3; 494/3; 579/3; 668/3; 709/3
 A:Note: D1007.15
 C:Superfamily: *Caenorhabditis elegans* hypothetical protein D1007.15

Query Match 1.1% Score 8; DB 2; Length 775;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 IPQSVSII 80
 |||||
 Db 262 IPQSVSII 269

RESULT 13

A40601
 ferripyoverdine receptor - *Pseudomonas aeruginosa*
 C:Species: *Pseudomonas aeruginosa*
 C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 29-Sep-1999
 C:Accession: A40601; S54000; F04227
 R;Poole, K.; Neshat, S.; Krebs, K.; Heinrichs, D.E.
 J. Bacteriol. 175, 4597-4604, 1993
 A:Title: Cloning and nucleotide sequence analysis of the ferripyoverdine receptor gene fli
 A:Reference number: A40601; MUID:9328663; PMID:8335619
 A:Accession: A40601
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-813 <POO>
 A:Cross-references: GB:L10210; NID:G294360; PIDN:AAA25819.1; PID:G294361
 R;Lamont, I.L.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S53998
 A:Accession: S54000

A:Molecule type: DNA
A:Residues: 1-715,'Y','717-744','YN','745-747','S','749-813 <LAN>
A:Cross-references: EMBL:U07559; NID:G1633044; PIDN:AA60199.1; PID:G466459
R:McMorran, B.V.; Merriman, M.E.; Rombel, I.T.; Lamont, I.L.
Gene 176, 55-59, 1996
A>Title: Characterisation of the pydE gene which is required for pyoverdine synthesis in
A:Accession: PC4227
A:Reference number: J05090; MUID:97075909; PMID:8918232
A:Molecule type: DNA
A:Residues: 1-49 <MCW>
A:Cross-references: GB:U07359
C:Genetics:
A:Gene: Fpva
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom
C:Keywords: membrane protein
F.184-317/Domain: tonB-dependent receptor carboxyl-terminal homology <TN>
F.538-813/Domain: tonB-dependent receptor carboxyl-terminal homology <TN>

Query Match 1.1%; Score 8; DB 2; Length 813;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 GAYVDLND 502
|||||
Db 574 GAYVDLND 581

RESULT 14
H83345
Ferripyoverdine receptor PA2398 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83345
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidis, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-815 <STO>
A:Cross-references: GB:AE004666; GB:AE004091; NID:G9948438; PIDN:AA605786.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: Fpva; PA2398
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal hom

Query Match 1.1%; Score 8; DB 2; Length 815;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 GAYVDLND 502
|||||
Db 574 GAYVDLND 581

RESULT 15
E87759
protein D1007.15 [imported] - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: E87759
R:anonymus, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A>Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog
A:Reference number: A75000; MUID:99069613; PMID:9851916
A>Note: see webites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C_elg
A:Accession: E87759
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-837 <STO>
A:Cross-references: GB:chr_1; PIDN:AA654232.1; PID:G2088812; GSPDB:GN00019; CESP:D1007.1
C:Genetics:
A:Gene: D1007.15
A:Map position: 1
C:Superfamily: Caenorhabditis elegans hypothetical protein D1007.15

Query Match 1.1%; Score 8; DB 2; Length 837;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 IPQSVSII 80
|||||
Db 324 IPQSVSII 331

RESULT 16
A89959
hypothetical protein SAI562 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: A89959
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A89959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1274 <KJR>
A:Cross-references: GB:BA000018; PID:G13701536; PIDN:BA642830.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SAI562

Query Match 1.1%; Score 8; DB 2; Length 1274;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 LAPYNGLP 268
|||||
Db 1003 LAPYNGLP 1010

RESULT 17
AD2109
hypothetical protein aer2427 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AD2109
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD2109
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-78 <KJR>
A:Cross-references: GB:BA000019; PIDN:BA674126.1; PID:G17131519; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: aer2427

Query Match 1.0%; Score 7; DB 2; Length 78;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 GLSLIGG 466
|||||

Db 26 GLSLIGG 32

RESULT 18
AC1886
hypothetical protein asr0636 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1886
R:Kansko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.; Nakazaki, N.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1886
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-84 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAF72594.1; PID:g17129982; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asr0636

Query Match 1.0%; Score 7; DB 2; Length 84;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 PASLREI 73
|||||
Db 40 PASLREI 46

RESULT 19
C81660
thioredoxin TC0826 [imported] - Chlamydia muridarum (strain Nigg)
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
C:Accession: C81660
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelsberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: C81660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-102 <TET>
A:Cross-references: GB:AE002349; GB:AE002160; NID:g7190851; PIDN:AAF39627.1; PID:g719085
A:Experimental source: strain Nigg (MoPn)
C:Genetics:
A:Gene: TC0826
C:Superfamily: thioredoxin; thioredoxin homology

Query Match 1.0%; Score 7; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 ASPRPAE 228
|||||
Db 57 ASPRPAE 63

RESULT 20
C97545
hypothetical protein AGR_C_2812 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: C97545
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194
A:Accession: C97545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87316.1; PID:g15156613; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2812
A:Map position: circular chromosome

Query Match 1.0%; Score 7; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLAA 20
|||||
Db 10 AATVLAA 16

RESULT 21
AD2764
conserved hypothetical protein Atul525 [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD2764
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD2764
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-108 <KUR>
A:Cross-references: GB:AB008688; PIDN:AAL42530.1; PID:g17739951; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul525
A:Map position: circular chromosome

Query Match 1.0%; Score 7; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLAA 20
|||||
Db 10 AATVLAA 16

RESULT 22
F87193
probable secreted protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: F87193
R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: F87193
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-112 <STO>
A:Cross-references: GB:AL450380; NID:g13093911; PIDN:CAC31790.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2274

Query Match 1.0%; Score 7; DB 2; Length 112;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 AATVLA 20
15 AATVLA 21
Db 15 AATVLA 21

RESULT 23
S29304
hyprochemical protein PA5055 [imported] - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 31-Dec-2000
C:Accession: S29304; B83013; S28376
R:Timm, A.; Steinhuechel, A.
Eur. J. Biochem. 209, 15-30, 1992
A:Title: Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus
A:Reference number: S29303; MUID:93011120; PMID:1396693
A:Accession: S29304
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-123 <TIM>
A:Cross-references: EMBL:X66592; NID:g45388; PIDN:CAA47150.1; PID:g45390
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Latbig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: B83013
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-123 <STO>
A:Cross-references: GB:AE004919; GB:AE004091; NID:g9951346; PIDN:AAG08440.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5055
C:Superfamily: Pseudomonas aeruginosa hypothetical protein 2 (phact 5' region)

Query Match 1.0%; Score 7; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 482 LHKASKT 488
9 LHKASKT 15
Db 9 LHKASKT 15

RESULT 24
T38756
60s ribosomal protein l32 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T38756
R:Olliver, K.; Harris, D.; Wood, V.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1996
A:Reference number: Z21809
A:Accession: T38756
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-127 <OLI>
A:Cross-references: EMBL:Z99296; PIDN:CAB16594.1; GSPDB:GN00066; SPDB:SPAC3H5.10
C:Genetics:
A:Gene: SPDB:SPAC3H5.10
A:Map position: 1
C:Superfamily: rat ribosomal protein l32

Query Match 1.0%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NNKTRY 575
56 NNKTRY 62
Db 56 NNKTRY 62

RESULT 25
T39562
60S ribosomal protein l32 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T39562
R:Pumelle, B.; Goffeau, A.; Wood, V.; Lyne, M.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998
A:Reference number: Z21863
A:Accession: T39562
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-127 <PUP>
A:Cross-references: EMBL:AL021767; PIDN:CAA16918.1; GSPDB:GN00067; SPDB:SPBC16C6.11
A:Experimental source: strain 972h-; cosmid c16C6
C:Genetics:
A:Gene: SPDB:SPBC16C6.11
A:Map position: 2
C:Superfamily: rat ribosomal protein l32

Query Match 1.0%; Score 7; DB 2; Length 127;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 569 NNKTRY 575
56 NNKTRY 62
Db 56 NNKTRY 62

RESULT 26
A84120
Atp synthase protein BH3761 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A84120
R:Takami, H.; Nakasone, K.; Takaki, Y.; Meno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A84120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-130 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BAE07480.1; GSPDB:GN001
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3761

Query Match 1.0%; Score 7; DB 2; Length 130;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 632 NLTMYQ 638
47 NLTMYQ 53
Db 47 NLTMYQ 53

RESULT 27
S71256
ribosomal protein l27a, cytosolic - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 20-Jun-2000
C:Accession: S71256
R:Greller, F.; Cooke, R.; Landie, M.; Raynal, M.; Delseny, M.
submitted to the EMBL Data Library, October 1995
A:Reference number: S71253
A:Accession: S71256
A:Molecule type: mRNA

A;Residues: 1-146 <GRE>
A;Cross-references: EMBL:X91959; NID:g1107486; PIDN:CAA63025.1; PID:g1107487
C;Superfamily: rat ribosomal protein L27a
C;Keywords: protein biosynthesis; ribosome

Query Match 1.0%; Score 7; DB 2; Length 146;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKVGMRY 318
|||||
Db 54 GKVGMRY 60

RESULT 28
C66732
transcription regulator [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 01-Mar-2002
C;Accession: C66732
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis se
A;Reference number: A66625; MUID:21235186; PMID:11337471
A;Accession: C66732
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-148 <STO>
A;Cross-references: GB:AE005176; PID:g1273787; PIDN:AAK04957.1; GSPDB:GN00146
A;Experimental source: strain IL1403
C;Genetics:
C;Superfamily: Escherichia coli arginine repressor argR

Query Match 1.0%; Score 7; DB 2; Length 148;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 364 KALAFDA 370
|||||
Db 77 KALAFDA 83

RESULT 29
R6BY29
ribosomal protein L27a.e, cytosolic - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein G3101; protein YGL103w; ribosomal protein YL29
C;Species: Saccharomyces cerevisiae
C;Date: 25-Feb-1985 #sequence_revision 30-Jun-1992 #text_change 21-Jul-2000
C;Accession: A02782; S45696; S64110
R;Kaeuffer, N.F.; Fried, H.M.; Schwindinger, W.F.; Jasin, M.; Warner, J.R.
Nucleic Acids Res. 11, 3123-3135, 1983
A;Title: Cycloheximide resistance in yeast: the gene and its protein.
A;Reference number: A02782; MUID:83220732; PMID:6304624
A;Accession: A02782
A;Molecule type: DNA
A;Residues: 1-149 <KAU>
A;Cross-references: EMBL:X01573; NID:g3619; PIDN:CAA25729.1; PID:g3620
A;Note: a mutation resulting in the replacement of Gln by Glu at position 38 confers res
R;Schwindinger, W.F.; Warner, J.R.
J. Biol. Chem. 262, 5690-5695, 1987
A;Title: Transcriptional elements of the yeast ribosomal protein gene CVH2.
A;Reference number: S45696; MUID:87194763; PMID:3553182
A;Accession: S45696
A;Molecule type: DNA
A;Residues: 1-16 <SCH>
A;Cross-references: EMBL:M19490; NID:g172481; PIDN:AAA35002.1; PID:g553140
R;Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
submitted to the Protein Sequence Database, May 1996
A;Reference number: S64071
A;Accession: S64110
A;Molecule type: DNA
A;Residues: 1-149 <RIE>

A;Cross-references: EMBL:Z72625; NID:g1322642; PIDN:CAA96808.1; PID:g1322643; GSPDB:GN00
A;Experimental source: strain S288C
C;Genetics:
A;Gene: SGD:CVH2; RPL29; MIPS:YGL103W
A;Cross-references: SGD:S0003071; MIPS:YGL103W
A;Map position: 7L
A;Introns: 17/1
C;Superfamily: rat ribosomal protein L27a
C;Keywords: protein biosynthesis; ribosome
F;2-149/Product: ribosomal protein L27a.e #status experimental <MAT>

Query Match 1.0%; Score 7; DB 1; Length 149;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKVGMRY 318
|||||
Db 54 GKVGMRY 60

RESULT 30
S77384
hypothetical protein sll1455 - Synecchocystis sp. (strain PCC 6803)
C;Species: Synecchocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C;Accession: S77384
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis
s.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S77384
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-155 <KAN>
A;Cross-references: EMBL:D90906; GB:AB001339; NID:g1652492; PIDN:BAA17487.1; PID:g165256
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Superfamily: Synecchocystis hypothetical protein sll1455

Query Match 1.0%; Score 7; DB 2; Length 155;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 272 NKKLPSL 278
|||||
Db 103 NKKLPSL 109

RESULT 31
AH3359
probable competence-damage protein [imported] - Brucella melitensis (strain 16W)
C;Species: Brucella melitensis
C;Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
R;DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A;Reference number: AD3252; PMID:11756688
A;Accession: AH3359
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-165 <KUR>
A;Cross-references: GB:AE008917; PIDN:AAL52043.1; PID:g17982810; GSPDB:GN00190
A;Experimental source: strain 16W
C;Genetics:
A;Gene: BMEI0862
A;Map position: I
C;Superfamily: Aquifex aeolicus conserved hypothetical protein aq_1996

Query Match 1.0%; Score 7; DB 2; Length 165;

Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 15 ATVLAAL 21
DB 149 ATVLAAL 155

RESULT 32

T10313
hypohectical protein 44 - Orygia pseudotsugata nuclear polyhedrosis virus
C:Species: Orygia pseudotsugata nuclear polyhedrosis virus, OpNPV
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
C:Accession: T10313
R:Ahrens, C.A.; Russell, R.R.; Funk, C.J.; Evans, J.; Harwood, S.; Rohmann, G.F.
Virology 229, 381-399, 1997
A:Title: The sequence of the Orygia pseudotsugata multinnucleocapsid nuclear polyhedrosis
A:Reference number: Z17011; MUID:97271300; PMID:9126251
A:Accession: T10313
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-166 <AHR>
A:Cross-references: EMBL:U75930; NID:g2934903; PIDN:AAC59043.1; PID:g1911290

Query Match 1.0%; Score 7; DB 2; Length 166;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
DB 105 AATVLA 111

RESULT 33

P83062
hypohectical protein PA4661 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: P83062
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lapid, K.; Lim,
J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: P83062
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-173 <STO>
A:Cross-references: GB:AE004880; GB:AE004091; NID:g9950912; PIDN:AAG08048.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4661

Query Match 1.0%; Score 7; DB 2; Length 173;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 VLAALSS 23
DB 9 VLAALSS 15

RESULT 34

C86231
hypohectical protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86231
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shum, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, J.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86231
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-174 <STO>
A:Cross-references: GB:AE005172; NID:g2160182; PIDN:AAB60745.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 1.0%; Score 7; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 DVGSGLN 206
DB 2 DVGSGLN 8

RESULT 35

PWECD
H+-transporting two-sector ATPase (EC 3.6.3.14) delta chain - Escherichia coli (strain K
N:Alternate names: hydrogen ion-transporting ATPase delta chain
C:Species: Escherichia coli
C:Date: 18-Dec-1981 #sequence_revision 02-Apr-1982 #text_change 01-Mar-2002
C:Accession: A93732; A90101; I41274; H65176; T45005; A01031
R:Gay, N.J.; Walker, J.E.
Nucleic Acids Res. 9, 3919-3926, 1981

A:Title: The *atp* operon: nucleotide sequence of the promoter and the genes for the membr
A:Reference number: A93732; MUID:82059437; PMID:6272190
A:Accession: A93732
A:Molecule type: DNA
A:Residues: 1-177 <GAY>
A:Cross-references: GB:U00264; GB:X00771; NID:g41023; PIDN:CAA23517.1; PID:g41028
R:Mabuchi, K.; Kanazawa, H.; Kayano, T.; Futai, M.
Biochem. Biophys. Res. Commun. 102, 172-179, 1981

A:Title: Nucleotide sequence of the gene coding for the delta subunit of proton-transloc
A:Reference number: A90101; MUID:82068433; PMID:6458296
A:Accession: A90101
A:Molecule type: DNA
A:Residues: 1-81, 'D', 83-177 <MAB>
A:Cross-references: GB:M12213; GB:M12213; NID:g145336; PIDN:AAA20044.1; PID:g145336
R:Kanazawa, H.; Futai, M.
Ann. N. Y. Acad. Sci. 402, 45-64, 1982

A:Title: Structure and function of H⁺-ATPase: What we have learned from *Escherichia coli*
A:Reference number: I41274; MUID:83176724; PMID:6301339
A:Accession: I41274
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-177 <RES>
A:Cross-references: GB:M25464; NID:g146318; PIDN:AAA83872.1; PID:g146322
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65176
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-177 <BLAT>
A:Cross-references: GB:AE000450; GB:U00096; NID:g1790166; PIDN:AAC76758.1; PID:g1790173;
R:Nielsen, J.; Hansen, F.G.; Hoppe, J.; Friedl, P.; Von Meyenburg, K.
Mol. Gen. Genet. 184, 33-39, 1981

A:Title: The nucleotide sequence of the *atp* gene coding for the F-0 subunits a, b, c an
A:Reference number: Z22893; MUID:82147764; PMID:6278247
A:Accession: T45005
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-177 <NIE>
A:Cross-references: EMBL:V00266; NID:g41031; PIDN:CAA23524.1; PID:g41034
C:Genetics:
A:Gene: atpH; uncH
A:Map position: 84 min
C:Complex: this is one of the five chains of the enzymatic component (F1) of the ATPase
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: ATP biosynthesis; hydrolase; membrane-associated complex

Query Match 1.0%; Score 7; DB 1; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 36
AC0954
A:Title: ATP synthase delta chain [imported] - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 17-May-2002
C:Accession: AC0954
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; PMID:11677608
A:Accession: AC0954
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03127.1; PID:g16504763; GSPDB:GN00176
C:Genetics:
A:Gene: STY3910
C:Superfamily: H⁺-transporting ATP synthase delta chain

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 37
F86059
H⁺-transporting two-sector ATPase delta chain [similarity] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 17-May-2002
C:Accession: F86059
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoudis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; PMID:21074935; PMID:11206551
A:Accession: F86059
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <STO>
A:Cross-references: GB:AB005174; NID:gl2518592; PIDN:AG58938.1; GSPDB:GN00145; UWGP:Z52
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: atpH
C:Superfamily: H⁺-transporting ATP synthase delta chain

Query Match 1.0%; Score 7; DB 2; Length 177;

Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 38
AH0500
H⁺-transporting two-sector ATPase (EC 3.6.3.14) delta chain [imported] - Yersinia pestis
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C:Accession: AH0500
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Hoiden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, S.;
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AH0500
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93573.1; PID:gl5982013; GSPDB:GN00175
C:Genetics:
A:Gene: atpH
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: hydrolase

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 39
E91213
H⁺-transporting two-sector ATPase (EC 3.6.3.14) delta chain [similarity] - Escherichia coli
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Jun-2002
C:Accession: E91213
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yagunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: E91213
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-177 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA038100.1; PID:gl3364152; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs4677
C:Superfamily: H⁺-transporting ATP synthase delta chain
C:Keywords: hydrolase

Query Match 1.0%; Score 7; DB 2; Length 177;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 208 DGSVRGR 214
|||||
Db 161 DGSVRGR 167

RESULT 40
D82490
Hypothetical protein VCA0195 [imported] - Vibrio cholerae (strain N16961 aerogroup O1)

C.Species: *Vibrio cholerae*
C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C.Accession: D82490
R.Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.; Chardon, D.; Ermojaeva, M.D.; Vamshayevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 4777-483, 2000
A.Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A.Reference number: A82035; MUID:20406833; PMID:10952301
A.Accession: D82490
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-182 <HE1>
A.Cross-references: GB:AE004359; GB:AE003853; NID:g9657575; PIDN:AAF96108.1; GSPDB:GN001
A.Experimental source: serogroup O1, strain N16961; biotype El Tor
C.Genetics:
A.Gene: VCA0195
A.Map position: 2

Query Match 1.0%; Score 7; DB 2; Length 182;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGCL 411
|||||
Db 107 LYARGCL 113

RESULT 41
T04031
hypothetical protein F17A8.180 - *Arabidopsis thaliana*
C.Species: *Arabidopsis thaliana* (mouse-ear cress)
C.Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Sep-1999
C.Accession: T04031
R.Bryan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999
A.Reference number: Z15184
A.Molecule type: DNA
A.Accession: T04031
A.Residues: 1-191 <BEV>
A.Cross-references: EMBL:AL049482
A.Experimental source: cultivar Columbia; BAC clone F17A8
C.Genetics:
A.Map position: 4
A.Introns: 53/3; 80/3
A.Note: F17A8.180
C.Superfamily: *Arabidopsis* hypothetical protein F17A8.180

Query Match 1.0%; Score 7; DB 2; Length 191;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 SLSQLYT 515
|||||
Db 29 SLSQLYT 35

RESULT 42
E95072
nitroreductase family protein [imported] - *Streptococcus pneumoniae* (strain TIGR4)
C.Species: *Streptococcus pneumoniae*
C.Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C.Accession: E95072
R.Tetzelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Held
on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Kadane, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A.Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.Title: Complete Genome Sequence of a virulent isolate of *Streptococcus pneumoniae*.
A.Reference number: A95000; MUID:21357209; PMID:11463916
A.Accession: E95072
A.Status: preliminary
A.Molecule type: DNA

A.Residues: 1-201 <KUR>
A.Cross-references: GB:AE005672; PIDN:AAK74774.1; PID:g14972098; GSPDB:GN00164; TIGR:SP4
A.Experimental source: strain TIGR4
C.Genetics:
A.Gene: SP0622

Query Match 1.0%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 ARVGGAN 703
|||||
Db 92 ARVGGAN 98

RESULT 43
B97940
nitroreductase (EC 1.-.-.-) [imported] - *Streptococcus pneumoniae* (strain R6)
C.Species: *Streptococcus pneumoniae*
C.Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C.Accession: B97940
R.Hoshing, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mahren, S.; P
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A.Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaaskunas, S.R.
A.Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
A.Reference number: A97872; MUID:21429245; PMID:11544234
A.Accession: B97940
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-201 <KUR>
A.Cross-references: GB:AE007317; PIDN:AAK99350.1; PID:g15458122; GSPDB:GN00174
C.Genetics:
A.Gene: nrd
C.Keywords: oxidoreductase

Query Match 1.0%; Score 7; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 697 ARVGGAN 703
|||||
Db 92 ARVGGAN 98

RESULT 44
B86874
hypothetical protein yuhE [imported] - *Lactococcus lactis* subsp. *lactis* (strain IL1403)
C.Species: *Lactococcus lactis* subsp. *lactis*
C.Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C.Accession: B86874
R.Bolotin, A.; Winkler, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrli
Genome Res. 11, 731-753, 2001
A.Title: The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ss
A.Reference number: A86625; MUID:21235186; PMID:11337471
A.Accession: B86874
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-208 <STO>
A.Cross-references: GB:AE005176; PID:g12725039; PIDN:AAK06092.1; GSPDB:GN00146
A.Experimental source: strain IL1403
C.Genetics:
A.Gene: yuhE

Query Match 1.0%; Score 7; DB 2; Length 208;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 645 IGGGVNA 651
|||||
Db 179 IGGGVNA 185

RESULT 45

AG2841
 outer membrane protein omp [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AG2841
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
 ; Karp, P.; Romero, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AG2841
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-209 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL43149.1; PID:gl7740625; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: omp
 A:Map position: circular chromosome

Query Match 1.0%; Score 7; DB 2; Length 209;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 209 GSVGRGV 215
 DB 117 GSVGRGV 123

Search completed: December 18, 2002, 06:59:29
 Job time : 40.351 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:53:55 ; Search time 10.2194 Seconds

(without alignments)
2930.303 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 722
Sequence: 1 MGQFMSVFRINMTATVLA.....NTFNIGSERWTANIRYSF 722

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	ID	Description
No.						
1	9	1.2	629	1	DXS_BACHD	Q9K871 bacillus ha
2	8	1.1	197	1	RNH2_PASMU	P57986 pasteurilla
3	8	1.1	542	1	ICP0_HSVB	P28990 equine herp
4	8	1.1	634	1	LEU1_MYCTU	P96420 mycobacteri
5	8	1.1	702	1	FOXA_SALTY	O56145 salmonella
6	8	1.1	720	1	FPTA_PSEAE	P42512 pseudomonas
7	8	1.1	815	1	FPVA_PSEAE	P48632 pseudomonas
8	7	1.0	102	1	THIO_CHLMU	Q9PJ13 chlamydia m
9	7	1.0	127	1	R32A_SCHPO	P79015 schizosacch
10	7	1.0	139	1	VG08_BPMU	O42335 schizosacch
11	7	1.0	146	1	R27B_ARATH	Q38480 bacteriopho
12	7	1.0	146	1	R27C_ARATH	O91433 arabidopsi
13	7	1.0	148	1	RL28_YEAST	P49637 arabidopsi
14	7	1.0	149	1	RL2A_BRYGR	P02406 saccharomy
15	7	1.0	166	1	Y024_NPVOP	P78987 erysipe gr
16	7	1.0	166	1	Y024_NPVOP	O10299 ergyia pseu
17	7	1.0	177	1	ATPD_ECOLI	P00831 escherichia
18	7	1.0	210	1	RGSB_CHICK	O9P840 gallus gall
19	7	1.0	210	1	RGSB_MOUSE	Q9G2B0 mus musculu
20	7	1.0	211	1	ROBP_RHLIV	Q52866 rhizobium 1
21	7	1.0	214	1	ET3_MOUSE	O48299 mus musculu
22	7	1.0	217	1	OMPV_VIBCH	P17266 vibrio chol
23	7	1.0	256	1	SPSA_BACSU	P39621 bacillus au
24	7	1.0	268	1	ISPE_AQUAE	O67060 aquifex aeo
25	7	1.0	286	1	TEHB_HAEIN	P45134 haemophilus
26	7	1.0	295	1	NADC_STNY3	P74301 haemophilus
27	7	1.0	318	1	FEPB_ECOLI	P14609 escherichia
28	7	1.0	333	1	TRPD_PASMU	P57856 pasteurilla
29	7	1.0	334	1	OTCC_HAEIN	P44770 haemophilus
30	7	1.0	344	1	YAAA_SCHPO	O09801 schizosacch
31	7	1.0	351	1	PHOE_CITFR	O01605 citrobacter
32	7	1.0	354	1	MURB_TREPA	O83128 treponema p
33	7	1.0	363	1	V363_ASFB7	P23164 african swi

34	7	1.0	382	1	IPAC_SHIDY	Q03946 shigella dy
35	7	1.0	382	1	IPAC_SHIFL	P18012 shigella fl
36	7	1.0	395	1	BN51_HUMAN	P05423 homo sapien
37	7	1.0	411	1	VGJG_HSVB	P28967 equine herp
38	7	1.0	411	1	VGJX_HSVB	P32514 equine herp
39	7	1.0	412	1	GLUP_BRUBA	O44623 brucella ad
40	7	1.0	427	1	Y118_TREPA	O83155 treponema p
41	7	1.0	431	1	POLC_HAEIN	P43175 haemophilus
42	7	1.0	443	1	UVSH_EMENT	O02398 emericella
43	7	1.0	454	1	SR54_AQUAE	O67615 aquifex aeo
44	7	1.0	456	1	EP1A_DICDI	P19624 dictyosteli
45	7	1.0	457	1	TBBI_PORPU	P50259 porphyra pu
46	7	1.0	477	1	XYNA_STRLI	P26514 streptomyce
47	7	1.0	485	1	CAT1_NICPL	P49315 nicotiana p
48	7	1.0	506	1	TCMO_PETCR	O43033 petroselinu
49	7	1.0	506	1	Y213_MYCPN	P75478 mycoplasma
50	7	1.0	520	1	C304_DROME	O9V917 drosophila
51	7	1.0	522	1	ACCD_SPTOL	O9M317 spinacia ol
52	7	1.0	554	1	HAP4_YEAST	P14064 saccharomy
53	7	1.0	557	1	C791_SORBI	O43135 sorghum bic
54	7	1.0	574	1	CO9_ONCMY	P06682 oncorhynch
55	7	1.0	586	1	CO9_FUGRU	P79755 fugu rubrip
56	7	1.0	632	1	YKGS_CABEL	P46555 caenorhabdi
57	7	1.0	632	1	PSO2_YEAST	P30620 saccharomy
58	7	1.0	661	1	RED1_RAT	P51400 rattus norv
59	7	1.0	714	1	Y1J2_YEAST	P47025 saccharomy
60	7	1.0	716	1	HEPA_HSVB	P28946 equine herp
61	7	1.0	741	1	RED1_HUMAN	P78563 homo sapien
62	7	1.0	747	1	GUND_CELFI	P50400 cellulomoc
63	7	1.0	759	1	Y363_METUA	O57809 methanococ
64	7	1.0	779	1	RHP9_SCHPO	P87074 schizosacch
65	7	1.0	808	1	PUBP_PSEPU	P38047 pseudomonas
66	7	1.0	811	1	YB03_SCHPO	P87145 schizosacch
67	7	1.0	819	1	PUPA_PSEPU	P25184 pseudomonas
68	7	1.0	823	1	SP97_YEAST	P38863 saccharomy
69	7	1.0	867	1	EP2_BLAHO	O17102 blastocysti
70	7	1.0	960	1	DLG1_DROME	P31007 dirosophila
71	7	1.0	963	1	UBP4_HUMAN	O13107 homo sapien
72	7	1.0	985	1	REVI_YEAST	P12689 saccharomy
73	7	1.0	1239	1	NME3_MOUSE	O01098 mus musculu
74	7	1.0	1284	1	NRK4_DROME	O94887 drosophila
75	7	1.0	1388	1	HRP3_SCHPO	O14139 schizosacch
76	7	1.0	1514	1	YD34_YEAST	O05471 saccharomy
77	7	1.0	1643	1	RPO0_NMV	P15095 narcissus m
78	7	1.0	1696	1	ITN2_HUMAN	O92436 rickettsia
79	7	1.0	2014	1	YUJ7_YEAST	O92436 rickettsia
80	7	1.0	2164	1	POLG_HRV89	O92436 rickettsia
81	6	0.8	44	1	RL34_RICCN	O92436 rickettsia
82	6	0.8	63	1	RL34_RICCP	O92436 rickettsia
83	6	0.8	63	1	YLM4_CABEL	P34407 caenorhabdi
84	6	0.8	65	1	YIME_BPHPI	P10437 bacteriopho
85	6	0.8	70	1	RL31_ECOLI	P02432 escherichia
86	6	0.8	72	1	PSBH_ARATH	P56780 arabidopsi
87	6	0.8	76	1	STBC_ECOLI	P11905 escherichia
88	6	0.8	82	1	YBDJ_ECOLI	P77506 escherichia
89	6	0.8	87	1	YOK2_CABEL	O11175 caenorhabdi
90	6	0.8	90	1	YET9_HAEIN	P44709 haemophilus
91	6	0.8	93	1	YQHV_BACSU	P49779 bacillus au
92	6	0.8	94	1	VP10_NPVAC	P04670 autocographa
93	6	0.8	100	1	REGN_BPP22	P04691 bacteriopho
94	6	0.8	101	1	Y12K_BNYVG	P19230 beet necrot
95	6	0.8	102	1	THIO_CHLTR	O84444 chlamydia t
96	6	0.8	103	1	H4_OXYNO	P18836 oxytricha n
97	6	0.8	103	1	MGE_RABIT	P47841 oryctolagus
98	6	0.8	103	1	MIR2_HUMAN	O9Y6H6 homo sapien
99	6	0.8	103	1	MIR2_MOUSE	O9W2H6 mus musculu
100	6	0.8	106	1	H4_EUPCR	P80739 euploies cr
101	6	0.8	107	1	MIR2_RAT	O9J1V7 rattus norv
102	6	0.8	109	1	YDGF_BACSU	P31494 bacillus au
103	6	0.8	110	1	CSCC_ECOLI	O26108 escherichia
104	6	0.8	111	1	Y518_METTH	O26108 methanobact
105	6	0.8	113	1	Y655_BACHD	O92436 rickettsia
106	6	0.8	119	1	RNPA_PASMU	P57915 pasteurilla

107	120	1	R22A_YEAST	P05749	saccharomyc	180	6	0.8	169	1	CRG2_RANTE	P02531	rana tempor
108	121	1	R22B_YEAST	P56628	saccharomyc	181	6	0.8	175	1	VS10_ROTBN	P08434	bovine rota
109	129	1	RS6E_ARCFU	O29739	archaeoglob	182	6	0.8	175	1	VS10_ROTBU	P04513	bovine rota
110	130	1	LYGE_MOUSE	Q64253	mus musculus	183	6	0.8	175	1	VS10_ROT8	P30032	human rotav
111	130	1	RR11_CVAPA	P48136	cyanophora	184	6	0.8	175	1	VS10_ROT8W	P03535	human rotav
112	130	1	RR11_WARPO	P06364	marchantia	185	6	0.8	176	1	VS10_MYCTU	Q06235	mycobacteri
113	130	1	RR11_MESVI	Q9muu9	mesocistima	186	6	0.8	177	1	NR13_COTJA	Q90343	coturnix co
114	130	1	RR11_ODOSI	P49499	odontella s	187	6	0.8	181	1	BCCP_ANASP	Q06881	anabaena ol
115	130	1	RR11_PINTH	P41630	pinus thunb	188	6	0.8	181	1	THIM_SPIOL	P07591	spinacia ol
116	130	1	RR11_PORPU	P51294	porphyra pu	189	6	0.8	182	1	MCRC_METVO	P11567	methanococc
117	130	1	RS11_SYNPG	O24709	synchococc	190	6	0.8	183	1	RL18_CICAR	O65729	cicer ariet
118	130	1	YERA_YEREN	P31490	yersinia en	191	6	0.8	183	1	RRP_DEIRA	Q59r82	deinococcus
119	130	1	YERA_YERPE	P31491	yersinia pe	192	6	0.8	183	1	TR13_HUMAN	Q15645	homo sapien
120	131	1	YGIT_ECOLI	Q46864	escherichia	193	6	0.8	183	1	YBCL_ECOLI	P77368	escherichia
121	133	1	ATPZ_BACFI	P22475	bacillus fi	194	6	0.8	186	1	WMA2_TRTV	P33494	turkey thin
122	133	1	EXD2_HELPY	O25987	helicobacte	195	6	0.8	187	1	RL18_ARATH	P42791	arabidopsis
123	133	1	YH35_SYNPG	Q31752	synchococc	196	6	0.8	187	1	RL18_HUMAN	Q07020	homo sapien
124	137	1	RS12_MYCPU	Q98gd6	mycoplasma	197	6	0.8	187	1	RL18_ICTPU	Q90yvo	ictalurus p
125	138	1	RR11_ARATH	P56802	arabidopsis	198	6	0.8	187	1	RL18_MOUSE	P35980	mus musculus
126	138	1	RR11_PEA	P06587	pisum sativ	199	6	0.8	187	1	RL18_OREMO	Q91836	oreochromis
127	138	1	RR11_SPIOL	P06506	spinacia ol	200	6	0.8	187	1	RL18_RAT	P12001	rattus norv
128	138	1	RR11_TOBAC	P06365	nicotiana t	201	6	0.8	187	1	RL18_XENLA	P02412	xenopus lae
129	138	1	VGAM_LAMB	P03702	bacterioph	202	6	0.8	187	1	RLNA_XENLA	P09897	xenopus lae
130	139	1	RBS_CVLN	P24683	cylindrothe	203	6	0.8	187	1	SURE_TREPA	O83434	treponema p
131	139	1	RBS_DETGO	O98946	odonula co	204	6	0.8	188	1	SLP_ECOLI	P37194	escherichia
132	139	1	RBS_ODOSI	P49521	odonella s	205	6	0.8	189	1	INAS_MOUSE	P07349	mus musculus
133	139	1	RBS_THANO	O98948	thalassiosi	206	6	0.8	190	1	INA9_MOUSE	P01573	mus musculus
134	141	1	CWAR_HUMAN	Q04762	homo sapien	207	6	0.8	190	1	INA9_MOUSE	P09235	mus musculus
135	141	1	Y085_ARCFU	Q30151	archaeoglob	208	6	0.8	192	1	Y56A_THEME	P58008	thermotoga
136	141	1	Y8OD_SULAC	P37820	sulfolobus	209	6	0.8	193	1	RL18_TRYBB	P50885	trypanosoma
137	142	1	FER_THEAC	P00218	thermoplas	210	6	0.8	195	1	PYRE_SULTO	Q970x1	sulfolobus
138	142	1	HEAC_HOPLI	R82315	hoplosternu	211	6	0.8	197	1	RUVA_BORBU	P70827	borrelia bu
139	142	1	YN77_RHIME	Q9x714	rhizobium m	212	6	0.8	198	1	SDH3_YEAST	P33421	borrelia bu
140	144	1	IL3_BOVIN	P49875	bos taurus	213	6	0.8	199	1	SYS_CRIGR	P26636	cricketulus
141	144	1	RL11_STGR	P36258	stretomyce	214	6	0.8	200	1	R8FA_ARATH	O65693	arabidopsis
142	144	1	SODM_BRAF1	P28761	branchiosto	215	6	0.8	203	1	CTFI_MOUSE	Q60753	mus musculus
143	145	1	OM25_HUMAN	P57105	homo sapien	216	6	0.8	203	1	CTFI_RAT	Q32086	rattus norv
144	145	1	RL2A_OSCB	O01358	oscheius br	217	6	0.8	203	1	RR22_HUMAN	Q92737	homo sapien
145	146	1	IL3_SHEEP	Q06435	ovis aries	218	6	0.8	204	1	YQF7_CABEL	Q09271	caenorhabdi
146	147	1	RL2A_MOUSE	P41415	mus musculus	219	6	0.8	206	1	OM25_RAT	Q9wv14	rattus norv
147	147	1	RL2A_RAT	P18445	rattus norv	220	6	0.8	207	1	UCRI_CHRVI	C31214	chromatium
148	147	1	RL2A_XENLA	P47830	xenopus lae	221	6	0.8	208	1	DSBA_SHIFL	P52235	shigella fl
149	148	1	CR20_XENLA	P52299	xenopus lae	222	6	0.8	208	1	UPP_ECOLI	P25532	escherichia
150	148	1	R28A_SCHPO	P36585	schizosacch	223	6	0.8	208	1	UPP_PASMU	Q9cpl8	pasteurella
151	148	1	R28B_SCHPO	P57728	schizosacch	224	6	0.8	209	1	IM23_MOUSE	Q9wt08	mus musculus
152	149	1	RL28_NEUCR	P08978	neurospora	225	6	0.8	209	1	IM23_RAT	Q35093	rattus norv
153	149	1	RL2A_DROME	P41092	drosophila	226	6	0.8	209	1	PYRE_COXBU	Q45918	coxiella bu
154	149	1	Y881_VIBCH	Q9ktm1	vibrioc chol	227	6	0.8	210	1	RHOD_HUMAN	Q00212	homo sapien
155	150	1	NAAA_BPT3	P20331	bacterioph	228	6	0.8	211	1	NUSB_ANASP	Q9yws6	anabaena sp
156	150	1	NAAA_BPT7	P00806	bacterioph	229	6	0.8	211	1	P1MT_PSEAE	P45683	pseudomonas
157	151	1	YOM1_PHOPR	P29741	photobacter	230	6	0.8	213	1	RB4A_HUMAN	P20338	homo sapien
158	151	1	YPR3_ECOLI	P10027	escherichia	231	6	0.8	213	1	RB4A_MOUSE	P56371	mus musculus
159	155	1	R57_BACST	P22744	bacillus st	232	6	0.8	213	1	RB4A_RAT	P05714	rattus norv
160	155	1	R57_BACSU	P21469	bacillus su	233	6	0.8	214	1	YQGE_BACST	P28754	bacillus st
161	155	1	R57_THETH	P17291	thermus the	234	6	0.8	216	1	KAD_FSEFU	Q9z409	pseudomonas
162	156	1	R57_BACHD	Q9z718	bacillus ha	235	6	0.8	217	1	HXB7_HUMAN	P09629	homo sapien
163	156	1	R57_LISMO	Q92714	listeria mo	236	6	0.8	218	1	CB25_TETTH	P09629	tetrahymena
164	156	1	R57_NEIMA	Q91r15	neisseria m	237	6	0.8	219	1	Y25K_BNYVF	P19229	beet necrot
165	156	1	R57_RALSO	O8xv09	raistonia s	238	6	0.8	221	1	NUSB_SVNY3	P74395	synchocyst
166	156	1	R57_THICU	O50564	thiobacillu	239	6	0.8	222	1	YMW7_YEAST	Q04272	saccharomyc
167	157	1	R57_EIKCO	P35642	eikenella c	240	6	0.8	222	1	HXC5_HUMAN	Q00444	homo sapien
168	158	1	N1FX_AZOVI	P34887	azotobacter	241	6	0.8	222	1	HXC5_MOUSE	P32043	mus musculus
169	159	1	COAD_SYNY3	Q55435	synchocyst	242	6	0.8	223	1	DEOC_PASMU	P57937	pasteurella
170	160	1	R57_RICCN	Q92j94	ricchetti	243	6	0.8	223	1	PHOP_ECOLI	P23836	escherichia
171	160	1	RS7_RICPR	P41081	rickettsia	244	6	0.8	224	1	PHOP_SALTY	P14146	salmonella
172	160	1	YSH2_CAEEL	Q09944	caenorhabdi	245	6	0.8	230	1	P1MT_ARATH	Q42539	arabidopsis
173	161	1	CHB2_BOMMO	P08828	bombyx mori	246	6	0.8	231	1	CUT2_CABEL	P34682	caenorhabdi
174	163	1	YC51_CVAPA	P48327	cyanophora	247	6	0.8	232	1	VSPA_BOTJA	P81661	bothrops ja
175	164	1	NOLB_RHIFR	P33208	rhizobium f	248	6	0.8	234	1	PUR7_SULTO	P09709	sulfolobus
176	164	1	NOLB_RHISN	P55713	rhizobium s	249	6	0.8	236	1	CD63_BOVIN	Q9x82	bos taurus
177	166	1	COAD_SYNP7	Q55235	synchococc	250	6	0.8	237	1	CD63_HUMAN	P08962	homo sapien
178	167	1	ET3_RAT	P13207	rattus norv	251	6	0.8	237	1	CD63_MOUSE	P41731	mus musculus
179	167	1	YCF3_GALSU	Q08814	galdieria s	252	6	0.8	237	1	CD63_RABIT	Q28709	oryctolagus

253	6	0.8	237	1	CD63_RAT	P26648	rattus novy	326	6	0.8	273	1	YF00_ECOLI	P53132	escherichia
254	6	0.8	239	1	CLDE_HUMAN	O95500	homo sapien	327	6	0.8	274	1	NHS_STRAP	P53391	streptomyce
255	6	0.8	239	1	CLDE_MOUSE	O95083	mus musculus	328	6	0.8	275	1	A464_CHVP1	O98514	paramycium
256	6	0.8	239	1	RS2_HAEN	P44371	mus musculus	329	6	0.8	276	1	OCCCT_AGRU	P35121	agrobacteri
257	6	0.8	240	1	GTO1_MOUSE	O09311	mus musculus	330	6	0.8	276	1	SLBP_DROME	O9yve6	drosophila
258	6	0.8	241	1	GTO1_HUMAN	P78417	homo sapien	331	6	0.8	277	1	TP2A_MERPE	O9yfe02	aeropyrum p
259	6	0.8	241	1	GTO1_PIG	O9n1f5	sus scrofa	332	6	0.8	278	1	PYRF_CLAPU	O9y1f8	cladosporiu
260	6	0.8	241	1	GTO1_RAT	O9z339	rattus novy	333	6	0.8	278	1	YTA_RHIS	P17986	rhizobium s
261	6	0.8	241	1	YH0H_STRGR	P24222	streptomyce	334	6	0.8	279	1	DAPE_AQUAE	O67693	aquifex aeo
262	6	0.8	242	1	LECT_PSOR	O2413	psophocarpu	335	6	0.8	280	1	KD0A_RHIME	O92699	rhizobium m
263	6	0.8	244	1	ARC3_GCBP	O00901	clostridium	336	6	0.8	280	1	S0M1_YEAST	P38858	saccharomyc
264	6	0.8	245	1	ICP3_HSV1N	P37319	herpes simp	337	6	0.8	280	1	THM1_YEAST	P53078	saccharomyc
265	6	0.8	245	1	X123_HUMAN	O15884	homo sapien	338	6	0.8	280	1	THM1_ECOS7	P58388	escherichia
266	6	0.8	245	1	Y0B2_HUMAN	P32688	escherichia	339	6	0.8	281	1	THM1_ECOLI	P31142	escherichia
267	6	0.8	246	1	RRM2_HUMAN	O9u143	homo sapien	340	6	0.8	281	1	Y645_PASMU	O9cm04	pasteurella
268	6	0.8	246	1	SURE_PASMU	P57955	pasteurella	341	6	0.8	282	1	CC08_CAEBL	O9cm04	caenorhabdi
269	6	0.8	246	1	MECG_SALTU	O8z397	salmomella	342	6	0.8	282	1	YARA_PROST	P46117	providencia
270	6	0.8	246	1	MECG_SALTU	P37457	salmomella	343	6	0.8	283	1	YRAL_HAEN	P45298	haemophilus
271	6	0.8	246	1	Y171_AQUAE	O66558	aquifex aeo	344	6	0.8	284	1	STAR_MESAU	P70114	mesocricetu
272	6	0.8	248	1	ARGB_IACPL	O08320	lactobacilli	345	6	0.8	286	1	DMA_HAEN	P44431	haemophilus
273	6	0.8	248	1	H162_CAMJB	O9pny5	campylobact	346	6	0.8	286	1	YMB1_CAEBL	P50093	caenorhabdi
274	6	0.8	248	1	ICP3_HSV1	P36313	herpes simp	347	6	0.8	288	1	ICMT_XENLA	O12947	x protein-s
275	6	0.8	250	1	ARC3_CLOM	O46134	clostridium	348	6	0.8	290	1	BUDR_DEIRA	P52666	klebsiella
276	6	0.8	250	1	DLX8_BRARE	O98879	brachydantio	349	6	0.8	290	1	IMP4_YEAST	P53941	saccharomyc
277	6	0.8	250	1	NGN2_HUMAN	O9z2a3	homo sapien	350	6	0.8	290	1	LEC_BAUPU	P16030	baumhina pu
278	6	0.8	250	1	SURE_VIBCH	O9ku19	vibriolo choi	351	6	0.8	290	1	YAS7_METUA	O58457	methanococc
279	6	0.8	251	1	Y095_HAEN	O57060	haemophilus	352	6	0.8	291	1	THIG_CYACA	O19915	cyanidium c
280	6	0.8	251	1	Y137_CHLPP	O9z246	chlamydia p	353	6	0.8	292	1	ARGB_DEIRA	O9ru66	deinococcus
281	6	0.8	252	1	ICP3_HSV1D	P37318	herpes simp	354	6	0.8	294	1	MENA_MYCLE	O07134	mycobacteri
282	6	0.8	252	1	PRGK_SALTU	P41786	salmomella	355	6	0.8	294	1	Y301_METUA	O57749	methanococc
283	6	0.8	252	1	RS2_PASMU	P57982	pasteurella	356	6	0.8	295	1	MURB_RICCN	O9z1t8	rickettsia
284	6	0.8	253	1	PSA3_SCHPO	O59770	schizosacch	357	6	0.8	295	1	MURB_RICCN	O9z1t8	rickettsia
285	6	0.8	253	1	SURE_ECOLI	P36664	escherichia	358	6	0.8	296	1	DAPE_MYCLE	P46814	mycobacteri
286	6	0.8	253	1	SURE_SALTU	O8x974	salmomella	359	6	0.8	296	1	MURB_CHLMU	O9p189	chlamydia m
287	6	0.8	253	1	Y883_MYCTU	O10545	mycobacteri	360	6	0.8	296	1	THTR_BOVIN	P00586	bos taurus
288	6	0.8	254	1	ATTA_TYRIN	P50725	trichoplusi	361	6	0.8	296	1	THTR_CRICR	P46635	criceulius
289	6	0.8	254	1	SURE_YERPE	O8z2p9	yerishina pe	362	6	0.8	296	1	THTR_HUMAN	O16762	homo sapien
290	6	0.8	255	1	S0L4_YEAST	P53115	saccharomyc	363	6	0.8	296	1	Y842_CAEBL	O09371	caenorhabdi
291	6	0.8	255	1	TH14_PYRHO	O59082	pyrococcus	364	6	0.8	297	1	BPHC_BURCE	P47228	buckholderi
292	6	0.8	255	1	VSPA_BOTAT	P04971	boehrops at	365	6	0.8	297	1	ISPA_BACST	O08291	baecillus st
293	6	0.8	256	1	CB4A_LYCPS	P27524	lycopersicio	366	6	0.8	298	1	YENK_CAEBL	O13408	caenorhabdi
294	6	0.8	256	1	TPIS_AGRTS	O8uey3	agrobacteri	367	6	0.8	299	1	ARGB_THETN	O87fco	thermonaer
295	6	0.8	257	1	Y453_AQUAE	O66760	aquifex aeo	368	6	0.8	300	1	NCSR_HUMAN	P00387	homo sapien
296	6	0.8	258	1	RPIA_YEAST	O12189	saccharomyc	369	6	0.8	300	1	P34_RICPR	O9zc05	rickettsia
297	6	0.8	258	1	SURE_COXBU	O9k121	coxijella bu	370	6	0.8	302	1	AROD_PYRAE	O8w59	pyrobaculum
298	6	0.8	258	1	Y256_MYCPN	P75421	mycoplasma	371	6	0.8	302	1	BPHC_PSEPS	P08695	pseudomonas
299	6	0.8	260	1	ER81_YEAST	P17261	saccharomyc	372	6	0.8	302	1	PPT2_HUMAN	O9umr5	homo sapien
300	6	0.8	261	1	YK83_CAEBL	P34349	caenorhabdi	373	6	0.8	302	1	PPT2_MOUSE	O38448	mus musculus
301	6	0.8	263	1	ICP3_HSV1P	P08853	herpes simp	374	6	0.8	302	1	PPT2_MOUSE	O70489	rattus novy
302	6	0.8	263	1	YG3S_YEAST	P27771	treponema p	375	6	0.8	303	1	URIC_MOUSE	P25688	mus musculus
303	6	0.8	263	1	YG3S_YEAST	P53389	saccharomyc	376	6	0.8	303	1	POCR_SALTU	O05587	salmomella
304	6	0.8	264	1	AMPW_BUCAL	P57324	buchnera ap	377	6	0.8	304	1	Y722_SYNYC	P50177	synecocyst
305	6	0.8	266	1	ETC1_STANU	O27657	methanobact	378	6	0.8	304	1	T212_LACIC	P58167	lactococcus
306	6	0.8	266	1	TH14_METH	O27657	methanobact	379	6	0.8	304	1	U049_HSVB	P28960	equine herp
307	6	0.8	267	1	COB2_HUMAN	P27701	homo sapien	380	6	0.8	305	1	ERA_VIBCH	O9Kp13	vibriolo choi
308	6	0.8	267	1	PM27_STRPU	O26616	strongyloce	381	6	0.8	307	1	CD5R_BOVIN	O28199	bos taurus
309	6	0.8	267	1	PROB_CLOAB	O97663	clostridium	382	6	0.8	307	1	CD5R_HUMAN	O15078	homo sapien
310	6	0.8	267	1	TRBJ_RHISN	P55400	rhizobium s	383	6	0.8	307	1	CD5R_MOUSE	O62938	mus musculus
311	6	0.8	267	1	Y123_THEMA	O9wax7	thermotoga	384	6	0.8	309	1	VP62_MRDV	P22120	maize rough
312	6	0.8	268	1	CLCR_HUMAN	O99895	homo sapien	385	6	0.8	310	1	UDP_HUMAN	O16831	homo sapien
313	6	0.8	268	1	FA12_RHIME	P58381	rhizobium m	386	6	0.8	311	1	FMT_MYCGE	P47605	mycoplasma
314	6	0.8	268	1	KNH1_YEAST	P50112	saccharomyc	387	6	0.8	311	1	SOX2_XENLA	O42569	xenopus lae
315	6	0.8	268	1	SURE_DEIRA	O9rt8	deinococcus	388	6	0.8	311	1	YAEK_ECOLI	P41409	escherichia
316	6	0.8	268	1	TRYP_STRGA	O54179	streptomyce	389	6	0.8	312	1	TRUB_BUCAL	P57456	buchnera ap
317	6	0.8	269	1	AQPI_HUMAN	P29972	homo sapien	390	6	0.8	312	1	VG06_BPMLS	O05278	mycobacteri
318	6	0.8	269	1	AQPI_MOUSE	O02013	mus musculus	391	6	0.8	313	1	ARAP_BACSU	P94529	bacillus su
319	6	0.8	269	1	AQPI_RAT	P29975	rattus novy	392	6	0.8	313	1	PENB_ECOLI	O02940	buckholderi
320	6	0.8	269	1	YFCA_ECOLI	P14008	escherichia	393	6	0.8	313	1	YDCU_ECOLI	P77156	escherichia
321	6	0.8	270	1	Y416_THEMA	O9wpy7	thermotoga	394	6	0.8	314	1	YG98_MYCTU	P48312	mycobacteri
322	6	0.8	271	1	AQPI_BOVIN	P47865	bos taurus	395	6	0.8	314	1	SOX2_CHICK	P58430	gallus gall
323	6	0.8	272	1	AQPI_SHEEP	P56401	ovis aries	396	6	0.8	316	1	Y874_CAEBL	O08618	caenorhabdi
324	6	0.8	273	1	YAEZ_YEAST	P29478	saccharomyc	397	6	0.8	317	1	SOX2_HUMAN	P48431	homo sapien
325	6	0.8	273	1	YAEZ_ECOLI	P77206	escherichia	398	6	0.8	317	1	YVCK_BACSU	O06974	bacillus su

399	6	0.8	318	1	TRXB_VIBCH	Q9kss4	vibrio chol	472	6	0.8	357	1	Z212_HUMAN	Q9udv6	homo sapien
400	6	0.8	320	1	ALDR_HORVU	P23901	hordeum vul	473	6	0.8	358	1	AZHS_CAVPO	O70159	cavia porce
401	6	0.8	320	1	MRAM_YERPE	Q8zif7	yersinia pe	474	6	0.8	358	1	CADH_MEDSA	P31656	medicago sa
402	6	0.8	320	1	SOX2_SHEEP	P54231	ovis aries	475	6	0.8	358	1	PHLC_TRYBB	P09194	trypanosoma
403	6	0.8	320	1	YD33_YEAST	Q12117	saccharomyc	476	6	0.8	358	1	RECA_XENBV	P96185	xenorhabdus
404	6	0.8	320	1	FCL3_HUMAN	Q13630	homo sapien	477	6	0.8	358	1	RECA_XENNE	P98505	xenorhabdus
405	6	0.8	321	1	TRYG_HUMAN	Q9nrr2	homo sapien	478	6	0.8	359	1	MAS2_AGRRH	P27873	agrobacteri
406	6	0.8	323	1	COBD_ARHPE	Q9yaa0	aeropyrum p	479	6	0.8	359	1	Y564_MYCLE	Q9ccn9	mycobacteri
407	6	0.8	324	1	YAMI_RHISN	P55568	rhizobium s	480	6	0.8	359	1	Y199_MYCTU	O07733	mycobacteri
408	6	0.8	325	1	CYF_SINP2	P26293	synecococc	481	6	0.8	361	1	GLNA_DAUCA	Q9axe3	caenor caro
409	6	0.8	325	1	RECA_PROVU	P26346	proteus vul	482	6	0.8	361	1	DCAM_PANAR	O04831	panulirus a
410	6	0.8	325	1	YBBG_ECOLI	P45394	escherichia	483	6	0.8	362	1	DCAM_IPOBA	Q9m6k1	ipomea bat
411	6	0.8	326	1	UNG2_HUMAN	P42674	homo sapien	484	6	0.8	362	1	DCAM_PHANI	Q96471	pharbitis n
412	6	0.8	326	1	VS09_ROTBA	Q00253	bovine rota	485	6	0.8	363	1	LACK_AGRRD	Q01937	agrobacteri
413	6	0.8	327	1	KDGT_ECO57	Q8x4q7	escherichia	486	6	0.8	363	1	VP43_NPVAC	P34050	autographa
414	6	0.8	327	1	KDGT_ECOLI	P32172	escherichia	487	6	0.8	363	1	YKL2_CAEEL	P42169	caenorhabdi
415	6	0.8	329	1	PIN2_CAEEL	Q19157	caenorhabdi	488	6	0.8	364	1	E2B1_PYPHO	O58433	pyrococcus
416	6	0.8	330	1	ASNA_HAEIN	P44338	haemophilus	489	6	0.8	364	1	MX_STRPY	P16946	streptococc
417	6	0.8	330	1	ASNA_STRPN	Q97nq0	streptococc	490	6	0.8	370	1	DNBJ_STRPE	P25048	streptomyce
418	6	0.8	330	1	ASNA_STRPY	Q99yu0	streptococc	491	6	0.8	370	1	TRZD_ENTCL	O87589	enterobacte
419	6	0.8	330	1	HBPB_ARATH	P43273	arabidopsis	492	6	0.8	372	1	M22_STRPY	P50469	streptococc
420	6	0.8	331	1	HXAI_MOUSE	P09022	mus musculus	493	6	0.8	373	1	CD14_BOVIN	Q95122	bos taurus
421	6	0.8	331	1	PUR7_ARCFU	O28996	archaeoglob	494	6	0.8	374	1	CARA_XYLFA	Q9pec2	xylella fas
422	6	0.8	331	1	PYRB_PSYTL	Q934t0	psychrobact	495	6	0.8	375	1	ADH4_KLULA	P49385	kluyveromyc
423	6	0.8	332	1	RUVB_STRPN	Q97sr6	streptococc	496	6	0.8	375	1	FLAB_VIBAN	Q56572	vibrio angu
424	6	0.8	333	1	HXAI_RAT	O08656	rattus norv	497	6	0.8	375	1	TLSX_MYCPN	P75159	mycoplasma
425	6	0.8	333	1	TRPD_HAEIN	P43858	haemophilus	498	6	0.8	375	1	YY10_MYCTU	Q49721	mycobacteri
426	6	0.8	334	1	MDHM_YEAST	P17505	saccharomyc	499	6	0.8	375	1	YV10_MYCTU	Q80716	mycobacteri
427	6	0.8	335	1	GPDA_HAEIN	P43798	haemophilus	500	6	0.8	376	1	HOSC_THETH	O87198	thermus the
428	6	0.8	335	1	HXAI_HUMAN	P49639	homo sapien	501	6	0.8	376	1	SPH_STRPY	P50470	streptococc
429	6	0.8	337	1	GPDA_PASMU	Q9cl17	pasteurella	502	6	0.8	379	1	FTSZ_MYCLE	Q9ccea	mycobacteri
430	6	0.8	339	1	KDGT_BRWCH	P15701	erwinia chr	503	6	0.8	379	1	FTSZ_MYCTU	O08378	mycobacteri
431	6	0.8	339	1	MOZL_CAEEL	Q9tzm2	caenorhabdi	504	6	0.8	379	1	VAT1_TORCA	P19333	torpedo cal
432	6	0.8	341	1	TRPD_LACCA	P17170	lactobacilli	505	6	0.8	381	1	RAPF_BAGSU	P17002	bacillus su
433	6	0.8	342	1	RECA_BRWCA	P26344	erwinia car	506	6	0.8	382	1	YC44_ODOSI	P49539	odontella s
434	6	0.8	342	1	YE22_MYCTU	P71691	mycobacteri	507	6	0.8	383	1	TLSX_MYCGE	Q49434	mycoplasma
435	6	0.8	342	1	YJ9J_YEAST	Q22306	saccharomyc	508	6	0.8	386	1	FTSZ_MYCKA	Q9khs5	mycobacteri
436	6	0.8	342	1	YWL2_CAEEL	Q22306	caenorhabdi	509	6	0.8	389	1	YDHP_ECO57	Q8x625	escherichia
437	6	0.8	343	1	NOV_XENLA	P51609	xenopus lae	510	6	0.8	389	1	YDHP_ECOLI	P77389	escherichia
438	6	0.8	343	1	POXA_BRUME	P58712	brucella me	511	6	0.8	391	1	NIFV_RHOSH	Q01181	rhodobacter
439	6	0.8	344	1	ALF_MYCTU	O06313	mycobacteri	512	6	0.8	392	1	SERA_MANSE	P14754	manduca sex
440	6	0.8	345	1	H1OM_BOVIN	P10950	bos taurus	513	6	0.8	392	1	YC4P_CAEEL	Q22915	caenorhabdi
441	6	0.8	346	1	GALM_ECOLI	P40681	escherichia	514	6	0.8	392	1	YFGI_ECOLI	P77774	escherichia
442	6	0.8	346	1	RFAF_HAEIN	P45042	haemophilus	515	6	0.8	393	1	DXR_RALSO	Q8x215	raistonia s
443	6	0.8	347	1	ABB2_MOUSE	P98199	mus musculus	516	6	0.8	394	1	BEXD_HAEIN	P22236	haemophilus
444	6	0.8	347	1	GBA5_DICDI	P34043	dictyosteli	517	6	0.8	394	1	MALA_BACST	Q45632	bacillus st
445	6	0.8	347	1	ID12_LACLA	Q9cif5	lactococcus	518	6	0.8	395	1	GP18_HUMAN	Q92643	homo sapien
446	6	0.8	347	1	Y999_MYCPN	P75593	mycoplasma	519	6	0.8	396	1	DXR_DEIFA	Q9ru84	deinococcus
447	6	0.8	347	1	YREG_HAEIN	P44720	haemophilus	520	6	0.8	396	1	IL3A_MOUSE	P26952	mus musculu
448	6	0.8	348	1	RECA_VIBAN	P26348	vibrio angu	521	6	0.8	397	1	GALI_TREPA	O83433	treponema p
449	6	0.8	349	1	TI29_ARATH	P82281	arabidopsis	522	6	0.8	398	1	CIW4_MOUSE	O88454	mus musculu
450	6	0.8	350	1	MLF1_MALFU	Q01940	malassezia	523	6	0.8	399	1	FTSZ_STRCO	P45500	streptomyce
451	6	0.8	350	1	XP22_HUMAN	O75695	homo sapien	524	6	0.8	400	1	ACHI_BOMMO	Q01383	bombyx mori
452	6	0.8	351	1	NOV_CHICK	P28686	gallus gall	525	6	0.8	400	1	CINA_ECOLI	P77808	escherichia
453	6	0.8	351	1	YAML_STRLM	P33823	streptomyce	526	6	0.8	401	1	YR07_CAEEL	Q09424	caenorhabdi
454	6	0.8	352	1	RECA_ECOLI	Q93017	escherichia	527	6	0.8	402	1	PGK_HELPJ	Q9zjpl	helicobacte
455	6	0.8	352	1	RECA_RALSO	Q8y1y6	raistonia s	528	6	0.8	402	1	PGK_HELPY	P56154	helicobacte
456	6	0.8	352	1	RECA_SALTY	Q8xet0	salmonella	529	6	0.8	403	1	S142_RAT	Q99m80	rattus norv
457	6	0.8	352	1	RECA_SHISO	Q9zff6	shigella so	530	6	0.8	405	1	PGK_CORGL	Q01655	corynebacte
458	6	0.8	353	1	LICH_PSEGL	O05490	pseudomonas	531	6	0.8	407	1	AATC_ORYSA	P27833	oryza sativ
459	6	0.8	353	1	NOV_COTJA	P42642	coturnix co	532	6	0.8	407	1	CPXD_AGRTU	P44667	agrobacteri
460	6	0.8	353	1	RECA_AERSA	P94190	aeromonas s	533	6	0.8	407	1	FTSZ_STRGR	P45501	streptomyce
461	6	0.8	353	1	RECA_ALCEU	Q9apb2	alcaligenes	534	6	0.8	408	1	PGK_SULSO	P50317	sulfolobus
462	6	0.8	353	1	RECA_ENTAG	P33037	enterobacte	535	6	0.8	408	1	YBX0_ARATH	Q9zq34	arabidopsis
463	6	0.8	353	1	RECA_SERWA	P17479	serriatia ma	536	6	0.8	409	1	IHH_XENLA	O91612	xenopus lae
464	6	0.8	354	1	FENR_CHLRE	P53991	chlamydomon	537	6	0.8	410	1	HXA3_HETFR	Q9ia21	heterodontu
465	6	0.8	354	1	RECA_HAEIN	P43705	haemophilus	538	6	0.8	412	1	ACDS_HUMAN	P16219	homo sapien
466	6	0.8	354	1	RECA_PROMI	P11406	proteus mir	539	6	0.8	412	1	ACDS_MOUSE	Q07417	mus musculu
467	6	0.8	354	1	RECA_VIBCH	P45383	vibrio chol	540	6	0.8	412	1	ACDS_RAT	P15651	rattus norv
468	6	0.8	355	1	EFTS_HELPJ	Q9zj71	helicobacte	541	6	0.8	414	1	YAFU_ECOLI	P04335	escherichia
469	6	0.8	355	1	EFTS_HELPY	P55975	helicobacte	542	6	0.8	414	1	YGJU_ECOLI	P42602	escherichia
470	6	0.8	355	1	RECA_YERPE	P37858	yersinia pe	543	6	0.8	415	1	RFBX_ECOLI	P37746	escherichia
471	6	0.8	357	1	AROB_STRPY	Q99yr3	streptococc	544	6	0.8	416	1	DADY_RHILO	Q98b75	rhizobium l

545	6	0.8	417	1	AGP_PROBE	Q52309	providencia	618	6	0.8	469	1	NIFN_RHISN	P55674	rhizobium s
546	6	0.8	417	1	PEGB_CRIFA	P08966	critihidia f	619	6	0.8	470	1	NIFD_METTM	Q50768	methanobact
547	6	0.8	417	1	PEGB_LEIMA	Q27663	leishmania	620	6	0.8	470	1	YICO_ECOLI	P1440	escherichia
548	6	0.8	417	1	PEGB_LEIME	Q27664	leishmania	621	6	0.8	471	1	GATA_THETH	Q91c3	thermus the
549	6	0.8	417	1	VGLD_HSVB	P24966	herp	622	6	0.8	471	1	VATS_DEIRA	Q9r87	deinococcus
550	6	0.8	417	1	VGLD_HSVB	Q08100	bovine herp	623	6	0.8	472	1	LEU2_BACHD	Q9K870	bachillus ha
551	6	0.8	419	1	PROA_STRKO	Q9rdk1	streptomyc	624	6	0.8	473	1	BGLB_MICBI	P98645	microbiopor
552	6	0.8	422	1	RASI_STRCN	Q02578	streptomyc	625	6	0.8	474	1	B1AR_FELIC	Q9t86	felis silve
553	6	0.8	424	1	COAA_BPFD	P03662	bacterioph	626	6	0.8	474	1	PEDB_LACTA	Q9cf3	lactococcus
554	6	0.8	424	1	COAA_BPM13	P03662	bacterioph	627	6	0.8	475	1	PPB_SERMA	P19147	serattia ma
555	6	0.8	424	1	IFPG_ARCFU	Q29663	archaeoglob	628	6	0.8	476	1	PPBH_PSEAE	P35453	pseudomonas
556	6	0.8	424	1	ZEP1_MOUSE	P08042	mus musculu	629	6	0.8	477	1	HRPB_PALSO	P31778	taletconia s
557	6	0.8	425	1	GCH2_MYCTU	P71684	mycobacteri	630	6	0.8	477	1	NIFD_METMP	P17526	methanococ
558	6	0.8	426	1	Y958_METYA	Q58368	methanococ	631	6	0.8	477	1	YGUJ_ECOLI	P25950	escherichia
559	6	0.8	427	1	DSL1_CHICK	P34822	gallus galli	632	6	0.8	478	1	GATA_MPCPN	P75534	mycoplasma
560	6	0.8	427	1	IRF3_HUMAN	Q14653	homo sapien	633	6	0.8	478	1	TBG_REPTI	P54405	reticulomyx
561	6	0.8	428	1	AMY1_ORYSA	P17654	oryza sativ	634	6	0.8	478	1	TTC_AQUOE	Q67338	aquiflex ae
562	6	0.8	428	1	GPBX_ORITIA	Q91178	oryzias lat	635	6	0.8	479	1	PEKC_LEIMA	P50312	leishmania
563	6	0.8	428	1	MYT8_THERM	P93749	thermus the	636	6	0.8	479	1	PEKC_LEIME	Q27665	leishmania
564	6	0.8	428	1	PURA_CLOAB	Q97487	clostridium	637	6	0.8	479	1	YBHI_HAEIN	P57048	haemophilus
565	6	0.8	430	1	BMP7_MOUSE	P23359	mus musculu	638	6	0.8	480	1	GLGI_RHIME	P18393	rhizobium m
566	6	0.8	432	1	PROA_DEIRA	Q97c19	deinococcus	639	6	0.8	480	1	NIFD_CVYAS	Q07662	cyanothec
567	6	0.8	432	1	TOLB_PSEAE	P50601	pseudomonas	640	6	0.8	480	1	YB9Q_YEAST	P38348	saccharomyc
568	6	0.8	432	1	YK27_CABEL	O16686	caenorhabdi	641	6	0.8	481	1	DDX6_XENLA	P54824	xenopus lae
569	6	0.8	432	1	YRKQ_BACSU	P54444	bacillus su	642	6	0.8	481	1	YAAU_HAEIN	P44555	haemophilus
570	6	0.8	434	1	YA47_XYLF	Q9pe11	xyella fas	643	6	0.8	482	1	CG23_SCHPO	P10815	schizosacch
571	6	0.8	437	1	INTR_SACCR	P22877	saccharopol	644	6	0.8	482	1	K6B2_HUMAN	Q9ub80	h ribosomal
572	6	0.8	439	1	GFO_ZYMMO	Q07982	zymomonas m	645	6	0.8	482	1	SAHH_STPAA	Q93666	streptomyc
573	6	0.8	439	1	PHT3_PSEPU	Q05183	pseudomonas	646	6	0.8	483	1	DDX6_HUMAN	P26186	homo sapien
574	6	0.8	439	1	SLAP_LACHE	P38059	lactobacill	647	6	0.8	483	1	DDX6_MOUSE	P54823	mus musculu
575	6	0.8	441	1	FAT6_RAT	P61537	rattus norv	648	6	0.8	483	1	KPYK_METEX	Q05118	methylobact
576	6	0.8	442	1	FTS2_CORGL	P94337	corynebacte	649	6	0.8	483	1	PREG_NETR	Q06712	neurospora
577	6	0.8	442	1	MCRY_METTH	P21111	methanobact	650	6	0.8	483	1	XYLB_KLEPN	P29444	klebsiella
578	6	0.8	443	1	YAAU_ECOLI	P31679	escherichia	651	6	0.8	485	1	K6B2_MOUSE	Q92184	mus musculu
579	6	0.8	444	1	GAT6_MOUSE	Q61169	mus musculu	652	6	0.8	485	1	NAB1_MESAS	Q35589	mesocricetu
580	6	0.8	444	1	GLGC_HAEIN	P43796	haemophilus	653	6	0.8	485	1	WDRD_HUMAN	Q9h14	homo sapien
581	6	0.8	444	1	PST1_YEAST	Q12355	saccharomyc	654	6	0.8	485	1	WDRO_MOUSE	Q91v09	mus musculu
582	6	0.8	444	1	TBB_ACHKL	P20802	achlya kleb	655	6	0.8	486	1	MURE_CHLMU	Q9PK66	chlamydia m
583	6	0.8	444	1	TBB_PHYCI	P05837	phytophhor	656	6	0.8	486	1	NAB1_HUMAN	Q13506	homo sapien
584	6	0.8	445	1	PERC_STRTR	Q56115	streptococc	657	6	0.8	486	1	RP54_VIBAN	Q08429	vibri
585	6	0.8	445	1	YIEG_ECOLI	P31466	escherichia	658	6	0.8	487	1	GLRK_AANPL	Q90218	anae plactyr
586	6	0.8	446	1	MOC_CHICK	P01875	gallus galli	659	6	0.8	487	1	Y442_MYCTU	P21611	mycobacteri
587	6	0.8	447	1	PME_MEDSA	Q42920	medicagu sa	660	6	0.8	492	1	CAT1_AARTH	Q96528	arabidopsi
588	6	0.8	448	1	FBL5_MOUSE	Q9wv19	mus musculu	661	6	0.8	492	1	CAT2_GOSHI	P30557	gossypium h
589	6	0.8	448	1	FBL5_RAT	Q9wv18	rattus norv	662	6	0.8	492	1	CAT2_LYCES	Q9xh33	lycopersico
590	6	0.8	449	1	GAT6_HUMAN	Q92908	homo sapien	663	6	0.8	492	1	CRT1_PANAN	P21685	pancoea ana
591	6	0.8	450	1	GSHR_STRTR	Q60151	streptococc	664	6	0.8	495	1	ACHP_RAT	P12352	rattus norv
592	6	0.8	450	1	UTH1_YEAST	P36133	saccharomyc	665	6	0.8	495	1	CBEP_COXBU	P19423	coxiella bu
593	6	0.8	451	1	CSSS_BACSU	Q31193	bacillus su	666	6	0.8	496	1	TM30_MOUSE	P15533	mus musculu
594	6	0.8	452	1	V51K_BPL79	P24125	lactococcus	667	6	0.8	497	1	NIFD_AANAS	P00474	anabaena sp
595	6	0.8	454	1	GUAD_MOUSE	Q97111	mus musculu	668	6	0.8	499	1	CIMS_HUMAN	Q95279	homo sapien
596	6	0.8	454	1	GUAD_RAT	Q9wtf6	rattus norv	669	6	0.8	499	1	RPB2_METVA	P41558	methanococ
597	6	0.8	454	1	PR11_SCHPO	O14215	schizosacch	670	6	0.8	500	1	NU4C_HORVU	P03307	hordium vul
598	6	0.8	455	1	A2AC_CAVPO	Q60476	cavia porce	671	6	0.8	500	1	NU4C_ORYSA	P12127	oryza sativ
599	6	0.8	455	1	PEKC_CRIFA	P08967	critihidia f	672	6	0.8	500	1	NU4C_WHEAT	P58420	triticism ae
600	6	0.8	455	1	SDHL_STRCO	O86564	streptomyc	673	6	0.8	501	1	ACHB_HUMAN	P11230	homo sapien
601	6	0.8	455	1	Y955_MYCTU	P15555	mycobacteri	674	6	0.8	502	1	K6B1_HUMAN	P23443	homo sapien
602	6	0.8	456	1	LEU2_STRAM	Q998j3	staphylococ	675	6	0.8	502	1	K6B1_RAT	P21425	rattus norv
603	6	0.8	456	1	LEU2_STRAM	P589j7	staphylococ	676	6	0.8	503	1	SIZ_BPSP	P54309	bacterioph
604	6	0.8	457	1	MUC_SUNMU	P20768	suncus muri	677	6	0.8	506	1	AER_ECOLI	P50466	escherichia
605	6	0.8	457	1	Y4YA_RHISN	P55709	rhizobium s	678	6	0.8	507	1	DHAL_MYCTU	O53773	mycobacteri
606	6	0.8	458	1	CBPN_HUMAN	P15169	homo sapien	679	6	0.8	508	1	HMOW_PIG	Q00274	sus scrofa
607	6	0.8	458	1	MEB1_DROME	P23128	drosophila	680	6	0.8	511	1	SVS_MOUSE	P26638	mus musculu
608	6	0.8	460	1	GNZ2_CLOJO	P37701	clostridium	681	6	0.8	511	1	XASA_ECOS7	P58229	escherichia
609	6	0.8	460	1	NIFN_RHILIO	Q98ap3	rhizobium 1	682	6	0.8	511	1	XASA_ECOLI	P39183	escherichia
610	6	0.8	462	1	COXX_YEAST	P21592	saccharomyc	683	6	0.8	512	1	XASA_SHIFL	O54152	shigella fl
611	6	0.8	462	1	YFHD_HAEIN	P44587	haemophilus	684	6	0.8	512	1	PER1_VOLCA	P11131	volvox cart
612	6	0.8	462	1	ZRAS_KLEOX	Q9ap80	klebsiella	685	6	0.8	513	1	RITI_YEAST	P23786	saccharomyc
613	6	0.8	463	1	D2DR_FUGUR	P53455	fugu rubrip	686	6	0.8	513	1	SVS_BOVIN	Q9gm8	bos taurus
614	6	0.8	464	1	GATH_YEAST	Q03557	saccharomyc	687	6	0.8	513	1	SVS_HUMAN	P49591	homo sapien
615	6	0.8	466	1	SAHH_BROME	Q9ye49	bruceella me	688	6	0.8	517	1	NCAP_SENDE	P04857	sendai vtru
616	6	0.8	466	1	SAHH_RHIME	Q92c1	rhizobium m	689	6	0.8	517	1	NCAP_SENDZ	P04858	sendai vtru
617	6	0.8	469	1	NIFD_METTH	Q27605	methanobact	690	6	0.8	517	1	VGLG_VSVYU	P04882	vesiculat s

691 1 ATPA_MYCGE
 692 1 ATPA_MYCPN
 693 1 SPKA_SYNY3
 694 1 FIXG_RHIME
 695 1 NCAP_PILHW
 696 1 NCAP_PILHW
 697 1 NCAP_SENDS
 698 1 NCAP_SENDF
 699 1 AROF_ARATH
 700 1 GOR_PANTR
 701 1 YJCC_ECOLI
 702 1 YCR2_ERWHE
 703 1 YOH1_CAEEL
 704 1 YPC1_CAEEL
 705 1 Y372_AQUAE
 706 1 SPG7_DICDI
 707 1 UD12_MOUSE
 708 1 P4H1_HUMAN
 709 1 ARSB_FELCA
 710 1 M24_STRPY
 711 1 HUP2_CHLKE
 712 1 CDY2_HUMAN
 713 1 NCAP_P12HT
 714 1 PGWU_AGRUT
 715 1 FAT2_YEAST
 716 1 CH63_RHIME
 717 1 CH63_BRAJA
 718 1 CHOD_STRSQ
 719 1 YTE4_CAEEL
 720 1 CH62_BRAJA
 721 1 FLIC_SHIFL
 722 1 AMYB_THETU
 723 1 FLKG_SALTY
 724 1 HIS5_YEAST
 725 1 LON2_BACSU
 726 1 HIS5_EMENT
 727 1 MAS2_YEAST
 728 1 MAS2_YEAST
 729 1 V12_HPV41
 730 1 ILV2_AERPE
 731 1 PKSJ_BACSU
 732 1 SYQ_HAEIN
 733 1 CSG_METJA
 734 1 FTHS_MOOTH
 735 1 SYO_PASMU
 736 1 INRI_BOVIN
 737 1 TREC_BACSU
 738 1 YC62_AQUAE
 739 1 PKNE_MYCTU
 740 1 TS13_MOUSE
 741 1 MCRA_METBA
 742 1 YAP3_YEAST
 743 1 Y559_MYCPN
 744 1 UI49_HCMVA
 745 1 C4GF_DROME
 746 1 MAO1_ECOLI
 747 1 ME22_SCHPO
 748 1 TERM_ADEGI
 749 1 PABP_YEAST
 750 1 PEX5_PICPA
 751 1 RECN_BACSU
 752 1 YTFM_HAEIN
 753 1 LLD1_KLUJA
 754 1 API_KLUJA
 755 1 MKS1_YEAST
 756 1 PPRR_DIDMA
 757 1 GF63_LEIDO
 758 1 GRN_HUMAN
 759 1 DCP2_SCHPO
 760 1 RCO3_NEUCR
 761 1 ACR2_NEUCR
 762 1 DPOL_BPMU5
 763 1 YE72_HUMAN

764 1 YA4C_SCHPO
 765 1 GP63_LEICH
 766 1 OPFA_LACLC
 767 1 OPFA_LACLC
 768 1 GP63_LEIMA
 769 1 RPB1_METTH
 770 1 RPB1_METTW
 771 1 YJGL_ECOLI
 772 1 GLMS_CHLMU
 773 1 GLMS_CHLTR
 774 1 DPOL_BPMU2
 775 1 LEUL_MYCLE
 776 1 PESC_SCHPO
 777 1 YD6C_SCHPO
 778 1 GLMS_HAEIN
 779 1 ENP4_MOUSE
 780 1 RPB1_METVA
 781 1 Y268_CHLMU
 782 1 ENP4_HUMAN
 783 1 HTPG_BORBU
 784 1 TERM_ADEB2
 785 1 YD54_MYCTU
 786 1 AMYG_NEUCR
 787 1 RSC4_YEAST
 788 1 GIDA_CIOAB
 789 1 UVRG_ANASP
 790 1 LU_HUMAN
 791 1 ESA8_TRYBB
 792 1 ESA8_TRYEQ
 793 1 GIAL_THETN
 794 1 GIAL_THETN
 795 1 Y561_HAEIN
 796 1 YKCA_CAEEL
 797 1 RPB1_METJA
 798 1 YHES_HAEIN
 799 1 GP63_LEIME
 800 1 NTP1_MSEPV
 801 1 METX_YEAST
 802 1 DD17_HUMAN
 803 1 YHJK_ECOLI
 804 1 FAS1_DROME
 805 1 DNK1_SYNP7
 806 1 NDC1_YEAST
 807 1 Y091_FOMPV
 808 1 YK79_MYCTU
 809 1 MTHR_YEAST
 810 1 ABD3_RAT
 811 1 VST2_HEYME
 812 1 LOX1_HUMAN
 813 1 FAS1_SCHAM
 814 1 TRB2_RHISN
 815 1 GR1A_BACCE
 816 1 EM70_YEAST
 817 1 VIDA_AGR5
 818 1 DNLS_SYNY3
 819 1 VINE_HUMAN
 820 1 VIDA_AGRRH
 821 1 KNS2_HUMAN
 822 1 MUTL_CLOPE
 823 1 HMUR_YERPE
 824 1 C1CL_RABIT
 825 1 KALM_HUMAN
 826 1 CYG2_RAT
 827 1 AKAB_RAT
 828 1 CICK_RABIT
 829 1 HEMR_YEREN
 830 1 INA_BACTL
 831 1 VID4_AGRTU
 832 1 YBP3_YEAST
 833 1 HELS_METTH
 834 1 PERO_DROME
 835 1 OAT5_HUMAN
 836 1 POL_SOCKV

Q99729 schizosacch
 P15706 leishmania
 Q07741 lactococcus
 Q09144 lactococcus
 P08148 leishmania
 Q07124 methanobact
 P09845 methanobact
 P39336 escherichia
 Q99144 c glucosami
 O84823 c glucosami
 O64235 mycobacteri
 O5cb76 mycobacteri
 O60164 schizosacch
 Q10324 schizosacch
 P44708 h glucosami
 Q9dbt4 mus musculu
 P11557 methanococc
 Q9pl40 chlamydia m
 Q9y227 homo sapien
 P42555 borrelia bu
 O55438 bovine aden
 P23799 trypanosoma
 P26337 trypanosoma
 Q8ac38 thermoanaer
 Q8r6k9 thermoanaer
 P44016 haemophilus
 P42083 caenorhabdi
 Q60181 methanococc
 P43150 leishmania
 Q9y439 melanoplus
 Q04533 saccharomyc
 Q92841 homo sapien
 P37649 escherichia
 P10674 drosophila
 P50020 synechococc
 P32500 saccharomyc
 O72896 fowlpox vir
 P46151 saccharomyc
 Q10687 mycobacteri
 P16970 rattus norv
 Q03500 hepatitis e
 P16050 homo sapien
 P10675 schistocerc
 P55399 rhizobium s
 O85467 bacillus ce
 P32802 saccharomyc
 P18594 agrobacteri
 P72588 synechocyst
 P13464 agrobacteri
 Q60504 homo sapien
 Q9bwl9 homo sapien
 Q9x186 clostridium
 Q56989 yersinia pe
 P51804 cryocollagus
 P23352 homo sapien
 P22717 rattus norv
 Q63014 rattus norv
 P51803 cryocollagus
 P31499 yersinia en
 P23382 bacillus th
 P09817 agrobacteri
 P38227 saccharomyc
 Q26901 methanobact
 Q01603 drosophila
 Q9y616 homo sapien
 P15629 soybean chl

837	6	0.8	693	1	THPA_HUMAN	P42166	homo sapien	910	6	0.8	821	1	PSGA_PYRHO	O57830	pyrococcus
838	6	0.8	701	1	LOXR_HUMAN	O75342	homo sapien	911	6	0.8	822	1	NFC1_PIG	O77638	sus scrofa
839	6	0.8	702	1	SYT_MYCLE	O07151	mycobacteri	912	6	0.8	823	1	NSF1_YEAST	P14907	saccharomy
840	6	0.8	709	1	SYT_ARATH	O04630	arabidopsis	913	6	0.8	826	1	CPRA_BACCH	O96597	bacillus th
841	6	0.8	710	1	GSOD_ERWCH	O01565	erwinia chr	914	6	0.8	826	1	PUHA_PSESP	O08017	pseudomonas
842	6	0.8	716	1	IKKE_HUMAN	O14164	homo sapien	915	6	0.8	826	1	RIR1_BBV	P03190	epstein-bar
843	6	0.8	716	1	RBP1_DHV11	P27153	dhori virus	916	6	0.8	829	1	RBPW_YEAST	P25328	saccharomy
844	6	0.8	717	1	IKKE_MOUSE	O91018	mus musculus	917	6	0.8	836	1	CRG1_HUMAN	O99490	homo sapien
845	6	0.8	719	1	FLBA_EMENI	P38093	emeritella	918	6	0.8	836	1	TGML_RABIT	P22758	oryctolagus
846	6	0.8	720	1	PSAB_EPHW	O9m1j8	ephedra twe	919	6	0.8	837	1	SECA_LISIN	O927Y3	listeria in
847	6	0.8	724	1	PCRA_BACST	P56255	bacillus st	920	6	0.8	837	1	SECA_LISMO	P47847	listeria mo
848	6	0.8	730	1	PIOD_CABEL	O20679	caenorhabdi	921	6	0.8	837	1	SM4G_MOUSE	O9wh7	mus musculus
849	6	0.8	732	1	ATZN_ECOLI	P37617	escherichia	922	6	0.8	838	1	PHS3_RAT	P53534	rattus norv
850	6	0.8	732	1	PRIA_ECOLI	P17888	escherichia	923	6	0.8	838	1	SM4G_HUMAN	O9ntn9	homo sapien
851	6	0.8	734	1	PSAB_ARATH	P56767	arabidopsis	924	6	0.8	840	1	UREA_CANNEN	P07374	canavalia e
852	6	0.8	734	1	PSAB_ORYSA	P12152	oryza sativ	925	6	0.8	841	1	SECA_BACCU	P28366	bacillus su
853	6	0.8	734	1	YKR2_CABEL	P34308	caenorhabdi	926	6	0.8	842	1	AXN_XENLA	O979Y0	xenopus lae
854	6	0.8	735	1	PSAB_MAIZE	P04967	zea mays (m	927	6	0.8	842	1	PHS2_BOVIN	P79334	bos taurus
855	6	0.8	737	1	PEN_DROME	O61345	drosophila	928	6	0.8	842	1	PHS2_HUMAN	P11217	homo sapien
856	6	0.8	738	1	TRFM_HUMAN	P08582	homo sapien	929	6	0.8	842	1	PHS2_MOUSE	O9wub3	mus musculus
857	6	0.8	740	1	PECI_PIG	O95242	sus scrofa	930	6	0.8	842	1	PHS2_RABIT	P00489	oryctolagus
858	6	0.8	741	1	CATA_ARCFU	O28050	archaeoglob	931	6	0.8	842	1	PHS2_RAT	P08812	rattus norv
859	6	0.8	747	1	VIVC_BPT7	P03725	bacterioph	932	6	0.8	842	1	PHS2_SHEEP	O18751	ovis aries
860	6	0.8	750	1	YKS7_YEAST	P34231	saccharomyc	933	6	0.8	843	1	PHS3_HUMAN	P11216	homo sapien
861	6	0.8	752	1	CLPE_STREP	P35394	streptococc	934	6	0.8	843	1	SECA_STAUV	O06446	staphylococ
862	6	0.8	755	1	ADIA_ECOLI	P28629	escherichia	935	6	0.8	844	1	APB2_YEAST	P32454	saccharomyc
863	6	0.8	758	1	PURL_CORAM	O9rtw9	corynebacte	936	6	0.8	844	1	PHSG_DROME	O9xt19	drosophila
864	6	0.8	759	1	ARY1_CALVI	P28513	calliphora	937	6	0.8	844	1	SECA_STACA	P47994	staphylococ
865	6	0.8	759	1	YEN1_YEAST	P40028	saccharomyc	938	6	0.8	847	1	PHS1_HUMAN	P06737	homo sapien
866	6	0.8	763	1	TSHR_BOVIN	O27987	bos taurus	939	6	0.8	850	1	PHS1_MOUSE	O96t01	mus musculus
867	6	0.8	764	1	TSHR_CANFA	P14763	canis famli	940	6	0.8	850	1	PHS1_RAT	P08901	rattus norv
868	6	0.8	764	1	TSHR_HUMAN	P16473	homo sapien	941	6	0.8	854	1	VGIB_RHOM6	P89053	rhesus cyto
869	6	0.8	764	1	TSHR_MOUSE	P47750	mus musculus	942	6	0.8	854	1	XRS2_YEAST	P33301	saccharomyc
870	6	0.8	764	1	TSHR_RAT	P21463	rattus norv	943	6	0.8	859	1	YD48_MYCTU	O11018	mycobacteri
871	6	0.8	764	1	TSHR_SHEEP	P56495	ovis aries	944	6	0.8	862	1	112S_HUMAN	O99665	homo sapien
872	6	0.8	765	1	Y008_HUMAN	O13398	homo sapien	945	6	0.8	865	1	PMU1_HUMAN	O43490	homo sapien
873	6	0.8	766	1	BCSB_SALTI	O82290	salmomella	946	6	0.8	865	1	SECA_HELPJ	O92157	helicobacte
874	6	0.8	766	1	BCSB_SALTY	O931n1	salmomella	947	6	0.8	865	1	SECA_HELPY	O25475	helicobacte
875	6	0.8	768	1	TRPG_EMENI	P06531	emeritella	948	6	0.8	868	1	MCW2_YEAST	P23469	saccharomyc
876	6	0.8	768	1	Y0DB_ECOLI	P25907	escherichia	949	6	0.8	868	1	NI80_YEAST	P33420	saccharomyc
877	6	0.8	771	1	PCRA_MYCTU	P71561	mycobacteri	950	6	0.8	869	1	CRAC_ECOLI	P25733	escherichia
878	6	0.8	776	1	KLPI_CHLRE	P46870	chlamydomon	951	6	0.8	872	1	DEOI_SULOH	O50607	sulfurispha
879	6	0.8	778	1	SEA2_MYCLE	O32922	mycobacteri	952	6	0.8	874	1	ATCL_MYCGB	P47317	mycoplasma
880	6	0.8	779	1	CAL1_BOVIN	P02453	bos taurus	953	6	0.8	875	1	SECA_BUCAL	P57297	buchnera ap
881	6	0.8	782	1	CHAO_TRICA	P82963	tricholium c	954	6	0.8	877	1	SECA_GUTHA	O78441	guillardia
882	6	0.8	782	1	ZAM_SYNY3	O46363	synecocyst	955	6	0.8	878	1	SECA_ANTSP	O05461	antitiamio
883	6	0.8	784	1	TIAR2_HUMAN	O68003	homo sapien	956	6	0.8	884	1	SECA_OLITU	O32743	olisthodisc
884	6	0.8	784	1	TIAR2_NACPA	O95053	macaca faec	957	6	0.8	884	1	SECA_PORPU	P51381	porphyra pu
885	6	0.8	785	1	IF16_HUMAN	O16666	homo sapien	958	6	0.8	886	1	OBP1_HAEIN	P45119	haemophilus
886	6	0.8	787	1	PFL_LACLA	O32797	lactococcus	959	6	0.8	887	1	OBP HEVER	P28947	equine herp
887	6	0.8	787	1	PFL_LACLC	O32799	lactococcus	960	6	0.8	888	1	SECA_ODOST	P46649	odontella s
888	6	0.8	787	1	XPKA_LACPE	O937f6	lactobacill	961	6	0.8	888	1	SYA_FRIILO	O9nng5	rhizobium l
889	6	0.8	790	1	DNLI_ARATH	O42572	arabidopsis	962	6	0.8	890	1	YOIN_ECOLI	P39838	escherichia
890	6	0.8	790	1	SUNI_MOUSE	O9d666	mus musculus	963	6	0.8	891	1	SECA_PAVLU	O15570	pavlova lut
891	6	0.8	790	1	SYFB_CHLMU	O9p1r1	chlamydia m	964	6	0.8	899	1	SECA_BORBU	O07497	borrella bu
892	6	0.8	790	1	SYFB_CHLTR	O84481	chlamydia t	965	6	0.8	900	1	YB75_YEAST	P38321	saccharomyc
893	6	0.8	792	1	YB5K_SSV1	P20210	sulfolobus	966	6	0.8	901	1	A180_MOUSE	O61548	mus musculus
894	6	0.8	795	1	Y210_HUMAN	O92609	homo sapien	967	6	0.8	901	1	CR14_MAIZE	O24585	zea mays (m
895	6	0.8	796	1	DHG_ECOLI	P15877	escherichia	968	6	0.8	901	1	DSC2_HUMAN	O02487	homo sapien
896	6	0.8	796	1	PHK_CLOAB	O97j63	clostridium	969	6	0.8	901	1	SECA_ECOLI	P14008	escherichia
897	6	0.8	797	1	SACB_STRMU	P11701	streptococc	970	6	0.8	901	1	SECA_HAEIN	P43803	haemophilus
898	6	0.8	798	1	YFR2_MOUSE	O9j1x3	mus musculus	971	6	0.8	902	1	NFC4_HUMAN	O14934	homo sapien
899	6	0.8	798	1	YCOO_MYCPN	O50088	mycoplasma	972	6	0.8	904	1	SECA_RHOCA	P52966	rhodospacter
900	6	0.8	806	1	SECA_MYCGB	P47318	mycoplasma	973	6	0.8	906	1	SECA_RICR	O9cxc7	rickettsia
901	6	0.8	808	1	SEA2_MYCTU	O50612	mycobacteri	974	6	0.8	912	1	PGCB_BOVIN	O28062	bos taurus
902	6	0.8	808	1	SECA_MYCPN	P75559	mycoplasma	975	6	0.8	915	1	A180_RAT	O05140	rattus norv
903	6	0.8	809	1	TORZ_ECOLI	P46823	escherichia	976	6	0.8	916	1	SECA_TREPA	O083394	treponema p
904	6	0.8	810	1	TIRA_ECOLI	O07736	escherichia	977	6	0.8	919	1	AMY_STRLI	O05884	streptomyce
905	6	0.8	812	1	FAED_ECOLI	P06970	escherichia	978	6	0.8	919	1	DNLI_HUMAN	P18858	homo sapien
906	6	0.8	817	1	PSBA_PYRRTU	P42850	pyrococcus	979	6	0.8	919	1	NOS3_RAT	O62600	rattus norv
907	6	0.8	819	1	PSBA_PYRAB	O92zh7	pyrococcus	980	6	0.8	919	1	RPO2_CAPVK	P16716	capripoxvir
908	6	0.8	820	1	GLGB_ORYSA	O01401	oryza sativ	981	6	0.8	922	1	YKFO_YEAST	P35736	saccharomyc
909	6	0.8	820	1	TRBE_AGRITU	P54910	agrobacteri	982	6	0.8	923	1	SECA_CAUCR	P38380	caulobacter

983 6 0.8 924 1 AACT DROME
984 6 0.8 926 1 SC24_YEAST
985 6 0.8 929 1 SYA_HALNI
986 6 0.8 932 1 SECA_SYN3
987 6 0.8 933 1 A8B2_HUMAN
988 6 0.8 934 1 Y321_MYCGE
989 6 0.8 937 1 ODOI_BACSU
990 6 0.8 940 1 SEAL_MYCLE
991 6 0.8 940 1 SECA_STRGR
992 6 0.8 943 1 DIP2_YEAST
993 6 0.8 945 1 A8B3_HUMAN
994 6 0.8 946 1 CITC_YEAST
995 6 0.8 947 1 SECA_STRCO
996 6 0.8 948 1 PMA5_ARATH
997 6 0.8 948 1 SECA_MYCBO
998 6 0.8 948 1 SECA_SYN7
999 6 0.8 949 1 SEAL_MYCTU
1000 6 0.8 956 1 NUT1_MAGGR

ALIGNMENTS

RESULT 1
DXS_BACHD STANDARD; PRT; 629 AA.
AC Q9K371;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1-deoxy-D-xylulose 5-phosphate synthase (EC 4.1.3.37) (1-
DE deoxyxylulose-5-phosphate synthase) (DXP synthase) (DXPS).
GN DXS OR BH2779.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RC MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirana C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Catalyzes the acyloln condensation reaction between C
CC atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield
CC 1-deoxy-D-xylulose-5-phosphate (DXP) (By similarity).
CC -!- CATALYTIC ACTIVITY: Pyruvate + D-glyceraldehyde 3-phosphate = 1-
CC deoxy-D-xylulose 5-phosphate + CO(2).
CC -!- COFACTOR: Thiamine pyrophosphate (By similarity).
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; first step.
CC -!- PATHWAY: Biosynthetic pathway to thiamine and pyridoxol; first
CC step.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TRANSKETOLASE FAMILY. DXS SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001516; BAB06498.1; -
CC InterPro; IPR000360; Transketolase.
CC Pfam; PF02779; transket_pyr; 1.
CC Pfam; PF02780; transketolase_C; 1.
CC TIGRFAMs; TIGR00204; dxs; 1.
CC PROSITE; PS00801; TRANSKETOLASE_1; 1.
CC PROSITE; PS00802; TRANSKETOLASE_2; 1.

KW Lyase; Flavoprotein; Thiamine pyrophosphate;
KW isoprene biosynthesis; Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 629 AA; 69279 MW; 7B5909B2F427C11E CRC64;
Query Match 1.2%; Score 9; DB 1; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.79; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Gaps 0;
Qy 470 GHYKIESGE 478
Db 299 GHYKIESGE 307
|||||
RESULT 2
RNH2_PASMU STANDARD; PRT; 197 AA.
AC P57986;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ribonuclease HII (EC 3.1.26.4) (RNase HII).
GN RNHB OR PM1998.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Pasteurella.
CC
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Pm70;
RC MEDLINE=21145866; PubMed=11248100;
RX May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC -!- FUNCTION: THIS ENZYME IS AN ENDONUCLEASE THAT DEGRADES THE RNA OF
CC RNA-DNA HYBRIDS SPECIFICALLY (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-
CC phosphomononucleosides.
CC -!- COFACTOR: MANGANESE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE RNASE HII FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE006236; AAK04082.1; -
CC HSSP; Q57599; 1EKE.
CC InterPro; IPR001352; RNase HII/HIII.
CC Pfam; PF01351; RNase_HII; 1.
CC Hydrolase; Nuclease; Endonuclease; Manganese; Complete proteome.
CC ACT_SITE 16 BY SIMILARITY.
CC ACT_SITE 108 BY SIMILARITY.
CC ACT_SITE 126 BY SIMILARITY.
CC SEQUENCE 197 AA; 21589 MW; CAA9F9C64442C7BB CRC64;
Query Match 1.1%; Score 8; DB 1; Length 197;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 360 EIKOKALA 367
Db 60 EIKOKALA 67
|||||
RESULT 3
ICPO_HSVB STANDARD; PRT; 532 AA.
ID_ICPO_HSVB
AC P28930;
DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Trans-acting transcriptional protein ICPO.
 GN 63.
 OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
 OC Viruses; deDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirinae.
 OK NCBI_TaxID=31520;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=92295566; PubMed=1318606;
 RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
 RT "The DNA sequence of equine herpesvirus-1";
 RL Virology 189:304-316(1992).
 RN [2]
 RN STRUCTURE BY NMR OF 1-63.
 RX MEDLINE=94087718; PubMed=8263911;
 RA Everett R.D., Barlow P.N., Milner A., Luisi B., Orr A., Hope G.,
 RA Lyon D.;
 RT "A novel arrangement of zinc-binding residues and secondary structure
 in the C3HC4 motif of an alpha herpes virus protein family.";
 RL J. Mol. Biol. 234:1038-1047(1993).
 RN [3]
 RN STRUCTURE BY NMR OF 1-63.
 RX MEDLINE=94172642; PubMed=8126734;
 RA Barlow P.N., Luisi B., Milner A., Elliott M., Everett R.D.;
 RT "Structure of the C3HC4 domain by 1H-nuclear magnetic resonance
 spectroscopy. A new structural class of zinc-finger.";
 RL J. Mol. Biol. 237:201-211(1994).
 CC -1- SIMILARITY: CONTRAINS 1 RING-TYPE ZINC FINGER.
 CC -1- SIMILARITY: TO OTHER HERPESVIRUSES ICPO PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, M86664; AAB02498.1; -
 DR PIR, I36801; WZBER5.
 DR PDB, 1CHC; 3O-APR-94.
 DR InterPro; IPR001841; Znf_ring.
 DR Pfam; PF00097; zf-C3HC4.1.
 DR SMART; SM00184; RING.1.
 DR PROSITE; PS00518; ZF_RING_1.
 DR PROSITE; PS50089; ZF_RING_2.1.
 DR Transcription regulation; Trans-acting factor; Activator; Zinc-finger;
 KM DNA-binding; 3D-structure.
 FT ZN FING 8 47
 FT METAL 8 47
 FT METAL 11 11
 FT METAL 24 24
 FT METAL 26 26
 FT METAL 29 29
 FT METAL 32 32
 FT METAL 33 43
 FT METAL 46 46
 FT METAL 46 46
 FT DOMAIN 210 217
 FT STRAND 19 21
 FT STRAND 22 25
 FT STRAND 26 28
 FT STRAND 30 31
 FT TURN 32 39
 FT HELIX 43 46
 FT TURN 43 46
 FT STRAND 53 54
 SQ SEQUENCE 532 AA; 58629 MW; B4CB7E16FA26FDFA CRC64;
 Query Match 1.1%; Score 8; DB 1; Length 532;
 Best Local Similarity 100.0%; Pred. No. 7.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ADVSGSLN 206
 Db 323 ADVSGSLN 330
 RESULT 4
 ID LEU1_MYCTU STANDARD; PRT; 644 AA.
 AC P96420; O69677;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 2-Isopropylmalate synthase (EC 4.1.3.12) (Alpha-isopropylmalate
 synthase) (Alpha-IPM synthetase).
 GN LEU1 OR RV3710 OR MT3813 OR MT025.058.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OK NCBI_TaxID=1773;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RA Chancham W., Namwat W., Palittapongarnpim P.;
 RT "Putative Mycobacterium tuberculosis alpha-isopropyl malate
 synthase gene."
 RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Davies R., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Hornsby T., Ungels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellton S., Squares S., Squares R.,
 RA Sulton J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishop W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains."
 RL Submitted (APR-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Catalyzes the condensation of the acetyl group of
 acetyl-CoA with 3-methyl-2-oxobutanoate (2-oxoisovalerate) to form
 3-carboxy-3-hydroxy-4-methylpentanoate (2-isopropylmalate).
 CC -1- CATALYTIC ACTIVITY: 3-carboxy-3-hydroxy-4-methylpentanoate + CoA =
 acetyl-CoA + 3-methyl-2-oxobutanoate + H(2)O.
 CC -1- PATHWAY: Leucine biosynthesis; first step.
 CC -1- SUBUNIT: Homotetramer (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE ALPHA-IPM SYNTHETASE / HOMOCITRATE
 SYNTHASE FAMILY. LEU1A 2 SUBFAMILY.
 CC -1- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
 in positions 49 and 77.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, U88526; AAB48096.1; ALT_FRAME.


```
DR EMBL; AL022121; CAA18032.1; --
DR EMBL; AE007178; AAK48181.1; ALT_INIT.
DR TIGR; MT3813; --
DR TubercuList; RV3710; --
DR InterPro; IPR002034; AIPM/Hcit_synth.
DR InterPro; IPR000891; HMGL-like.
DR Pfam; PF00682; HMGL-like; 1.
DR TIGRFAMs; TIGR00970; leuA_yeast; 1.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00816; AIPM_HOMOCIT_SYNTH_2; 1.
KW Leucine biosynthesis; Lyase; Complete proteome.
FT DOMAIN 577 614 INSERT.
FT CONFLICT 628 633 TTASUR -> NRPA (IN REF. 1).
FT CONFLICT 642 644 AAR -> RHARTALN (IN REF. 1).
SQ SEQUENCE 644 AA; 70113 MW; 881A9AAEA7F8FB71 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 644;
Best Local Similarity 100.0%; Pred. No. 8.8; Mismatches 0; Indels 0; Gaps 0;
Matches 8; Conservative 0;

Qy 585 GVETETISG 592
Db 522 GVETETISG 529
|||||

RESULT 5
FOXA_SALTY
ID FOXA_SALTY STANDARD; PRT; 702 AA.
AC Q56145.
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferrioxamine B receptor precursor.
GN FOXA OR STM0364.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McLelland M., Sanderson K.E., Sprieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2."
RL Nature 413:852-856(2001).
[2]
RN SEQUENCE OF 1-374 FROM N.A.
RC STRAIN=SL1344;
RX MEDLINE=95203118; PubMed=10103258;
RA Kingsley R.A., Reissbrodt R., Rabsch W., Ketley J.M., Tsolis R.M.,
RA Everest P., Dougan G., Baumber A.J., Roberts M., Williams P.H.;
RT "Ferrioxamine-mediated iron(III) utilization by Salmonella
enterica."
RL Appl. Environ. Microbiol. 65:1610-1618(1999).
[3]
RN SEQUENCE OF 1-244 FROM N.A.
RC STRAIN=ATCC 14028;
RX MEDLINE=95370139; PubMed=7642488;
RA Tsolis R.M., Baumber A.J., Stojiljkovic I., Hefron F.;
RT "Fur regulon of Salmonella typhimurium: identification of new iron-
regulated genes."
RL J. Bacteriol. 177:4628-4637(1995).
CC -!- FUNCTION: FERROXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
TONB PROTEIN (BY SIMILARITY). MAY PLAY A ROLE IN INTESTINAL
COLONIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE008712; AAL19318.1; --
DR EMBL; AF060876; AAC15464.1; --
DR EMBL; U62282; AAB04552.1; --
DR HSSP; P06971; 1BY5
DR StyGene; SG10646; foxA.
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB_box; Complete proteome.
FT SIGNAL 1 30 POTENTIAL.
FT CHAIN 31 702 FERROXAMINE B RECEPTOR.
FT CONFLICT 246 249 MAYP -> SPYL (IN REF. 2).
FT CONFLICT 255 255 G -> R (IN REF. 2).
FT CONFLICT 279 279 F -> S (IN REF. 2).
FT CONFLICT 352 352 A -> P (IN REF. 2).
FT CONFLICT 368 374 HRLVGI -> SSSAGGD (IN REF. 2).
SQ SEQUENCE 702 AA; 77685 MW; 7E2CB405B09E860D CRC64;

Query Match 1.1%; Score 8; DB 1; Length 702;
Best Local Similarity 100.0%; Pred. No. 9.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 74 POSVSIIT 81
Db 60 POSVSIIT 67
|||||

RESULT 6
FPTA_PSEAE
ID FPTA_PSEAE STANDARD; PRT; 720 AA.
AC P42512;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fe(III)-pyochelin receptor precursor.
GN FPTA OR PA4221.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas
OC NCBI_TaxID=287;
[1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.
RC STRAIN=PAO / IA602;
RX MEDLINE=94117363; PubMed=8288523;
RA Ankenbauer R.G.; Quan H.N.;
RT "FptA, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a
phenolate siderophore receptor homologous to hydroxamate siderophore
receptors."
RL J. Bacteriol. 176:307-319(1994).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -!- FUNCTION: HIGH-AFFINITY OUTER MEMBRANE RECEPTOR REQUIRED FOR THE
TRANSPORT OF FE(III)-PYOCHELIN.
-----
```



```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U03161; AAC43213.1; -.
CC EMBL: AE004839; AAG07609.1; -.
CC InterPro: IPR000531; TonB_BoxC.
CC Pfam: PF00593; TonB_BoxC.1.
CC PROSITE: PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB DEPENDENT REC_2; 1.
CC Outer membrane; Iron transport; Transport; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 38
CC CHAIN 39 720 FE(III)-PYOCHELIN RECEPTOR.
CC SITE 703 720 TONB C-TERMINAL BOX.
CC SEQUENCE 720 AA; 79992 MW; DA796313116E0CC2 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 720;
Best Local Similarity 100.0%; Pred. No. 9.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 170 VNLVRRKP 177
DB 176 VNLVRRKP 183

RESULT 7
PPVA_PSEAE STANDARD; PRT; 815 AA.
AC P48632; Q51339;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferripyoverdine receptor precursor.
GN PPVA OR PA2398.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.
RC STRAIN=CD10;
RX MEDLINE=93328663; PubMed=8335619;
RA Poola K., Neshat S., Krebs K., Heinrichs D.E.;
RT "Cloning and nucleotide sequence analysis of the ferripyoverdine
RT receptor gene fpyv of Pseudomonas aeruginosa."
RL J. Bacteriol. 175:4597-4604(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltzy U., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Polger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reiser J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.

```

```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
CC STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L10210; AAA25819.1; -.
CC EMBL: U07359; AAB60199.1; -.
CC EMBL: AE004666; AAG05786.1; -.
CC PIR: A40601; A40601.
CC InterPro: IPR000531; TonB_BoxC.
CC Pfam: PF00593; TonB_BoxC.1.
CC PROSITE: PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB DEPENDENT REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB_Box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 43
CC CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
CC SITE 798 815 TONB C-TERMINAL BOX.
CC CONFLICT 716 716 Y -> F (IN REF. 1).
CC CONFLICT 745 746 MISSING (IN REF. 1).
CC CONFLICT 750 750 S -> R (IN REF. 1).
CC SEQUENCE 815 AA; 91167 MW; A11AE4AA5A290F35 CRC64;

Query Match 1.1%; Score 8; DB 1; Length 815;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 495 GAVYDLND 502
DB 574 GAVYDLND 581

RESULT 8
THIO_CHLMU STANDARD; PRT; 102 AA.
AC Q9PJF3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thioiodoxin (TRX).
GN TRXA OR TC0826.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Mopn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uetcheback T., Berry K., Bass S.,
RA Linner K., Weidman J., Khoult H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- FUNCTION: Participates in various redox reactions through the
CC reversible oxidation of its active center dithiol to a disulfide
CC and catalyzes dithiol-disulfide exchange reactions.
CC -1- SIMILARITY: BELONGS TO THE THIOREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```


CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC -----
 CC EMBL; AE002349; AAF39627.1; --
 DR HSP; F80579; IQW.
 DR TIGR; TC0826; --
 DR InterPro; IPR000063; ThioRed.
 DR Pfam; PF00085; thioRed; 1.
 DR PRINTS; PR00421; THIOREDOXIN.
 DR TIGRPFAMs; TIGR01068; thioRedoxin; 1.
 DR PROSITE; PS00194; THIOREDOXIN; 1.
 KW Redox-active center; Electron transport; Complete proteome.
 FT DISULFID 28 31 REDOX-ACTIVE (BY SIMILARITY).
 SQ SEQUENCE 102 AA; 11147 MW; C171B646D393428C CRC64;

Query Match 1.0%; Score 7; DB 1; Length 102;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 222 ASPRAE 228
 |||||
 DB 57 ASPRAE 63

RESULT 9

ID R32A SCHPO STANDARD; PRT; 127 AA.
 AC P79015;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L32-A.
 GN RPL32A OR SPAC3H5.10.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]_
 RN SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomycetes pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 4-127 FROM N.A.
 RA Kawamukai M.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- MISCELLANEOUS: THERE ARE TWO GENES FOR L32 IN S.POMBE.
 CC -!- SIMILARITY: BELONGS TO THE L32E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

CC -----
 CC EMBL; Z99296; CAB16594.1; --
 DR EMBL; AB000914; BAA19212.1; --
 DR InterPro; IPR001515; Ribosomal L32E.
 DR Pfam; PF01655; Ribosomal L32e; 1.
 DR ProDom; PD003823; Ribosomal L32E; 1.
 DR PROSITE; PS00580; RIBOSOMAL_L32E; 1.
 KW Ribosomal protein; Multigene family.
 SQ SEQUENCE 127 AA; 14458 MW; 4BBB3CEB2E359E05 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 127;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NKKTRY 575
 |||||
 DB 56 NKKTRY 62

RESULT 10

ID R32B SCHPO STANDARD; PRT; 127 AA.
 AC O42935;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 60S ribosomal protein L32-B.
 GN RPL32B OR SPBC16C.11.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 [1]_
 RN SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volkart G., Aert R., Robben J., Grymonprez B.,
 RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomycetes pombe."
 RL Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 4-127 FROM N.A.
 RA Kawamukai M.;

RL Nature 415:871-880 (2002).
 CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L32 IN S.POMBE.
 CC -1- SIMILARITY: BELONGS TO THE L32E FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AL021767; CAA16918.1; -
 CC InterPro; IPR001515; Ribosomal_L32E.
 CC Pfam; PF01655; Ribosomal_L32e; 1.
 CC PROSITE; PD003823; Ribosomal_L32e; 1.
 CC DR PROSITE; PS00580; RIBOSOMAL_L32E; FALSE_NEG.
 CC Ribosomal protein, Multigene family.
 CC SEQUENCE 127 AA; 14445 MW; 0149CA472053669B CRC64;

Query Match 1.0%; Score 7; DB 1; Length 127;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 569 NNKTRY 575
 Db 56 NNKTRY 62

RESULT 11

VG08_BPMU STANDARD; PRT; 139 AA.

AC Q38480;

DT 16-OCT-2001 (Rel. 40; Created)

DT 16-OCT-2001 (Rel. 40; Last sequence update)

DT 16-OCT-2001 (Rel. 40; Last annotation update)

DE Hypothetical protein gp8 (Protein E7).

GN 8 OR E7.

OS Bacteriophage Mu.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;

OC Mu-like viruses.

OX NCBI_TaxId=10677;

RN [1]

RP SEQUENCE FROM N.A.

RA Pries H., Brauer B., Schmidt C., Kamp D.;

RT "Sequence of the left end of Mu.;"

RL (In) Symonds N., Tounssaint A., van de Putte P., Howe M.M. (eds.);

RL Phase Mu, pp.277-296, Cold Spring Harbor Laboratory Press,

RL New York (1987).

RL [2]

RP SEQUENCE FROM N.A.

RA Morgan G., Hatfull G., Hendrix R.;

RT "Genome of bacteriophage Mu and comparison with the Haemophilus

RT influenzae Mu-like prophage Flumu.;"

RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL, M64097; AAA32391.1; -

DR EMBL, AF083977; AAF01134.1; -

KW Hypothetical protein.

SEQUENCE 139 AA; 16310 MW; 80ACAC9B18F10A93 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 139;

Best Local Similarity 100.0%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 66 IPASIRE 72
 Db 129 IPASIRE 135

RESULT 12

R27B_ARATH STANDARD; PRT; 146 AA.

ID R27B_ARATH

AC Q9LR33;

DT 15-JUN-2002 (Rel. 41; Created)

DT 15-JUN-2002 (Rel. 41; Last sequence update)

DT 15-JUN-2002 (Rel. 41; Last annotation update)

DE 60S ribosomal protein L27a-B.

GN RPL27AB OR ATG623290 OR F26F24.13 OR F26F24_23.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucosids II; Brassicales; Brassicaceae; Arabidopsie.

OX NCBI_TaxId=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,

RA Utecherback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

RT thaliana."

RL Nature 408:816-820 (2000).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;

RT "Riken Arabidopsis full length cDNA clones (RAFLs) sequenced by the

RT SSP consortium (Salk/Stanford/PGEC)."

RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 3-141 FROM N.A.

RC STRAIN=cv. Columbia;

RA Hofte H.;

RT Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL, AC005292; AAF68998.1; -

DR EMBL, AF349525; AAK15572.1; -

DR EMBL, AF324716; AAG40067.1; -

DR EMBL, AF410280; AAK95266.1; -

DR EMBL, 226208; CAA81190.1; -

DR InterPro; IPR001196; Ribosomal_L15.

DR Pfam; PF00256; L15; 1.

DR PROSITE; PS00475; RIBOSOMAL_L15; FALSE_NEG.

KW Ribosomal protein.


```
FT CONFLICT 3 4 TA -> DG (IN REF. 3).
FT CONFLICT 31 31 G -> V (IN REF. 3).
FT CONFLICT 139 139 G -> C (IN REF. 3).
SQ SEQUENCE 146 AA; 16292 MW; 0CA6C019D8B1A565 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 GKVGMRY 318
DB 54 GKVGMRY 60

RESULT 13
R27C ARATH
ID R27C ARATH STANDARD; PRT; 146 AA.
AC P49637.
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 60S ribosomal protein L27A-C.
GN RPL27AC OR ATIG70600 OR F24J13.17 OR F5A18.22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=97336304; PubMed=9193080;
RA Cooke R., Raynal M., Laudie M., Delseny M.;
RT "Identification of members of gene families in Arabidopsis thaliana
RT by contig construction from partial cDNA sequences: 106 genes
RT encoding 60 cytoplasmic ribosomal proteins.";
RL Plant J. 11:1127-1140(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.B., Krenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Utecherback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE OF 1-87 FROM N.A.
RC STRAIN=cv. Columbia;
RA Raynal M., Grellet F., Laudie M., Meyer Y., Cooke R., Delseny M.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; X01573; CAA25729.1; -
CC EMBL; M19490; AAA35002.1; -
CC EMBL; Z72625; CAA96808.1; -

or send an email to license@isb-sib.ch.
CC
CC EMBL; X91959; CAA63025.1; -
CC EMBL; AC011663; AAG52338.1; -
CC EMBL; Z17767; CAA79059.1; -
CC InterPro: IPR001196; Ribosomal_L15.
CC Pfam: PF00256; L15.1
CC PROSITE: PS00475; RIBOSOMAL_L15; FALSE_NEG.
KW Ribosomal protein.
SQ SEQUENCE 146 AA; 16455 MW; B04CB9121A0EE94C CRC64;

Query Match 1.0%; Score 7; DB 1; Length 146;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 312 GKVGMRY 318
DB 54 GKVGMRY 60

RESULT 14
RL28 YEAST
ID RL28 YEAST STANDARD; PRT; 148 AA.
AC P02406;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L28 (L27A) (L29) (YL24) (RP62).
GN RPL28 OR CYH2 OR YGL103W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83220732; PubMed=6304624;
RA Kaufer N.F., Fried H.M., Schwindinger W.F., Jasin M., Warner J.R.;
RT "Cycloheximide resistance in yeast: the gene and its protein.";
RL Nucleic Acids Res. 11:3123-3135(1983).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97435481; PubMed=9290212;
RA Rieger M., Brueckner M., Schaefer M., Mueller-Auer S.;
RT "Sequence analysis of 203 kilobases from Saccharomyces cerevisiae
RT chromosome VII.";
RL Yeast 13:1077-1090(1997).
RN [3]
RP SEQUENCE OF 1-15 FROM N.A.
RX MEDLINE=87194763; PubMed=3553182;
RA Schwindinger W.F., Warner J.R.;
RT "Transcriptional elements of the yeast ribosomal protein gene CYH2.";
RL J. Biol. Chem. 262:5690-5695(1987).
RN [4]
RP NUCLEAR LOCALIZATION SIGNAL.
RX MEDLINE=90107966; PubMed=2104804;
RA Underwood M.R., Fried H.M.;
RT "Characterization of nuclear localizing sequences derived from yeast
RT ribosomal protein L29.";
RL EMBO J. 9:91-99(1990).
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL; X01573; CAA25729.1; -
CC EMBL; M19490; AAA35002.1; -
CC EMBL; Z72625; CAA96808.1; -
```



```

DR EMBL; K01162; AAA35003.1; -.
DR PIR; A02782; R6B129.
DR SGD; S0003071; RPL28.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein; Cycloheximide resistance.
FT INIT_MET 0 0
FT DOMAIN 6 12 NUCLEAR LOCALIZATION SIGNAL.
FT DOMAIN 23 29 NUCLEAR LOCALIZATION SIGNAL.
FT VARIANT 37 37 E -> Q (CONFERES RESISTANCE TO
CYCLOHEXIMIDE, AN INHIBITOR OF
POLYPEPTIDE ELONGATION).
SQ SEQUENCE 148 AA; 16591 MW; F3E1CD53C6292FAA CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 148;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKGVMRY 318
Db 53 GKGVMRY 59

```

```

RESULT 15
RL2A_ERYGR STANDARD; PRT; 149 AA.
AC P78987;
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S ribosomal protein L27a (L29).
OS Erythrae gramini (subsp. hordei) (Grass mildew).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Erythrales; Erysiphaceae; Blumeria.
OX NCBI_TaxID=62688;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CR3;
RA Justesen A.F.; Christiansen S.K.; Giese H.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE L15P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y11394; CAA72204.1; -.
DR InterPro; IPR001196; Ribosomal_L15.
DR Pfam; PF00256; L15; 1.
DR PROSITE; PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein.
SQ SEQUENCE 149 AA; 16786 MW; 599E3BD3195171 CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 149;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 312 GKGVMRY 318
Db 54 GKGVMRY 60

RESULT 16
Y024_NPYOP STANDARD; PRT; 166 AA.
AC O10259;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)

```

```

DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 18.4 kDa protein (ORF44).
OS Oryza pseudotsugata multicausid polyhedrosis virus (OpMVNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Oryza pseudotsugata multicausid polyhedrosis virus genome."
RT polyhedrosis virus genome."
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U75930; AAC59043.1; -.
KW Hypothetical protein.
SQ SEQUENCE 166 AA; 18390 MW; 1D6C0F2FAF877DDF CRC64;

```

```

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 166;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AATVLA 20
Db 105 AATVLA 111

```

```

RESULT 17
ATPD_ECOLI STANDARD; PRT; 177 AA.
AC P00831;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ATP synthase delta chain (EC 3.6.3.14).
GN ATP OR UNCH OR PAPE OR B3735 OR Z5233 OR ECS4677.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85121806; PubMed=6395859;
RA Walker J.E., Gay N.J., Saraste M., Eberle A.N.;
RT "DNA sequence around the Escherichia coli unc operon. Completion of
RT the sequence of a 17 kilobase segment containing asna, oric, unc,
RT glms and phos."
RL Biochem. J. 224:799-815(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=82059437; PubMed=6272190;
RA Gay N.J., Walker J.E.;
RT "The atp operon: nucleotide sequence of the promoter and the genes
RT for the membrane proteins, and the delta subunit of Escherichia coli
RT ATP-synthase."
RL Nucleic Acids Res. 9:3919-3926(1981).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=82147764; PubMed=6278247;
RA Nielsen J., Hansen F.G., Hoppe J., Friedl P., von Meyenburg K.;
RT "The nucleotide sequence of the atp genes coding for the F0 subunits
RT a, b, c and the F1 subunit delta of the membrane bound ATP synthase

```



```
ET3_MOUSE          STANDARD;          PRT;    214 AA.
ID   ET3_MOUSE
AC   P48299;
DT   01-FEB-1996 (Rel. 33, Created)
DI   01-FEB-1996 (Rel. 33, Last sequence update)
DR   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Endothelin-3 precursor (ET-3).
GN   EDN3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   MEDLINE=95094302; PubMed=8001160;
RX   Baynash A.G., Hosoda K., Gaid A., Richardson J.A., Emoto N.,
RA   Hammer R.E., Yanagisawa M.;
RT   "Interaction of endothelin-3 with endothelin-B receptor is essential
RT   for development of epidermal melanocytes and enteric neurons.";
RL   Cell 79:1277-1285(1994).
CC   -!- FUNCTION: ENDOTHELINS ARE ENDOTHELIUM-DERIVED VASOCONSTRICTOR
CC   PEPTIDES.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE ENDOTHELIN/SARAFOTOXIN FAMILY.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; U32330; AAB60509.1; -.
DR   HSP; P05305; 1EDN.
DR   MGD; MGI:95285; Edn3.
DR   InterPro; IPR003641; Endothelin.
DR   InterPro; IPR001928; Endothelin.tox.
DR   Pfam; PF00322; endothelin; 1.
DR   PRINTS; PR00365; ENDOTHELIN.
DR   ProDom; PD005286; Endothelin; 1.
DR   SMART; SM00272; END; 2.
DR   PROSITE; PS00270; ENDOTHELIN; 2.
KW   Cleavage on pair of basic residues; Vasoconstrictor; Multigene family;
Signal.
FT   SIGNAL          1..25   POTENTIAL.
FT   PEPTIDE         97..117 ENDOTHELIN-3.
FT   DOMAIN          159..173 ENDOTHELIN LIKE.
FT   DISULFID        97..111 BY SIMILARITY.
FT   DISULFID        99..107 BY SIMILARITY.
SQ   SEQUENCE 214 AA; 23322 MW; 8911260872D6A713 CRC64;
Query Match          1.0%; Score 7; DB 1; Length 214;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  348 ADDKACA 354
DB  164 ADDKACA 170
|||||
RESULT 22
OMPW_VIBCH
ID   OMPW_VIBCH          STANDARD;          PRT;    217 AA.
AC   P17266; O9KLR0;
DT   01-AUG-1990 (Rel. 15, Created)
DI   16-OCT-2001 (Rel. 40, Last sequence update)
DR   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Outer membrane protein W precursor.
GN   OMPW OR VCA0867.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
SPSAL_BACSU
ID   SPSAL_BACSU          STANDARD;          PRT;    256 AA.
AC   P39621;
DT   01-FEB-1995 (Rel. 31, Created)
DI   01-FEB-1995 (Rel. 31, Last sequence update)
DR   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Spore coat polysaccharide biosynthesis protein spsa.
GN   SPSA OR IPA-63D.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90245674; PubMed=2336399;
RA   Jalajakumari M.B., Manning P.A.;
RT   "Nucleotide sequence of the gene, ompW, encoding a 22kDa immunogenic
RT   outer membrane protein of Vibrio cholerae.";
RL   Nucleic Acids Res. 18:2180-2180(1990).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   STRAIN=El Tor N16961 / Serotype O1;
RC   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA   McDaniel L., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   Emdonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
CC   -!- SUBCELLULAR LOCATION: Outer membrane.
CC   -!- SIMILARITY: TO E.COLI OMPW AND P.OLEOVORANS ALKL.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X51948; CAA36210.1; -.
DR   EMBL; AE004414; AAF96765.1; -.
DR   PIR; S09509; S09509.
DR   TIGR; VCA0867; -.
KW   Signal; Outer membrane; Antigen; Complete proteome.
FT   SIGNAL          1..21
FT   CHAIN           22..217
FT   CONFLICT        7..7   MISSING (IN REF. 1).
FT   CONFLICT        15..16 SS -> LA (IN REF. 1).
FT   CONFLICT        59..59 Q -> H (IN REF. 1).
FT   CONFLICT        79..79 A -> R (IN REF. 1).
FT   CONFLICT        122..122 F -> N (IN REF. 1).
FT   CONFLICT        142..142 G -> S (IN REF. 1).
FT   CONFLICT        147..148 AG -> NA (IN REF. 1).
FT   CONFLICT        179..179 S -> Y (IN REF. 1).
FT   CONFLICT        210..210 M -> I (IN REF. 1).
SQ   SEQUENCE 217 AA; 23316 MW; BF376BB11EF2797B CRC64;
Query Match          1.0%; Score 7; DB 1; Length 217;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  17 VLAALSS 23
DB  10 VLAALSS 16
|||||
RESULT 23
SPSAL_BACSU
ID   SPSAL_BACSU          STANDARD;          PRT;    256 AA.
AC   P39621;
DT   01-FEB-1995 (Rel. 31, Created)
DI   01-FEB-1995 (Rel. 31, Last sequence update)
DR   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Spore coat polysaccharide biosynthesis protein spsa.
GN   SPSA OR IPA-63D.
OS   Bacillus subtilis.
OC   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX   NCBI_TaxID=1423;
RN   [1]
RP   SEQUENCE FROM N.A.
```


RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
 RA Presecan E., Sarcana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=96044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherst S.,
 RA Borries R., Bourrier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Erington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holstappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Katamata D., Kasahara Y., Klaerr-Bianchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portelle D., Porwoll G., Prescott A.M.,
 RA Pirescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeder R., Scoffone F.,
 RA Setiguchi J., Sekowska A., Serron S.J., Serron P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tanakoshi A., Tanaka T., Terpestra P., Tognoni K.,
 RA Tosiato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viati A., Wamburt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winiers P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RX MEDLINE=99280694; PubMed=10350455;
 RA Charneck S.J., Davies G.J.;
 RT "Structure of the nucleotide-diphospho-sugar transferase, Spgs from
 Bacillus subtilis, in native and nucleotide-complexed forms.";
 RL Biochemistry 38:6380-6385(1999).
 CC -!- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
 CC SPORE COAT.
 CC -!- PATHWAY: Spore coat polysaccharide biosynthesis.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, X73124; CA51619.1; -;
 DR EMBL, 299123; CAB15817.1; -;
 DR PDB, 1QG8; 21-APR-00.
 DR PDB, 1QGO; 04-MAY-00.
 DR PDB, 1QGS; 04-MAY-00.
 DR Subtilast; BG10609; spsa.
 DR InterPro; IPR001173; Glycosyltransferase_2.
 DR Pfam; PF00535; Glycosyltransferase; 3D-structure; Complete proteome.
 KW

FT DISULFID 155 243
 SQ SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;
 Query Match
 Best Local Similarity 1.0%; Score 7; DB 1; Length 256;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 572 KTRVAA 578
 |||||
 DB 74 KTRVAA 80
 RESULT 24
 ID ISPE_AQUAE STANDARD; PRT; 268 AA.
 AC 067060;
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 41, Last annotation update)
 DE 4-diphosphocytidylyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148) (CMK
 (4-cytidine-5'-diphospho)-2-C-methyl-D-erythritol kinase).
 GN ISPE OR AQ.915.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VF5;
 RX MEDLINE=98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjaj M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358(1998).
 CC -!- FUNCTION: CATALYZES THE PHOSPHORYLATION OF THE POSITION 2 HYDROXY
 CC GROUP OF 4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL (BY
 CC SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + 4-(cytidine 5'-diphospho)-2-C-methyl-D-
 CC erythritol = ADP + 2-phospho-4-(cytidine 5'-diphospho)-2-C-methyl-
 CC D-erythritol.
 CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; fourth
 CC step.
 CC -!- SIMILARITY: BELONGS TO THE ISPE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL, AE000713; AAC07027.1; -;
 DR InterPro; IPR001745; GHMPkinse_ATP.
 DR InterPro; IPR004424; ISPE.
 DR Pfam; PF00288; GHMP_kinases; 1.
 DR TIGRFAMs; TIGR00154; ispe; 1.
 KW Transferase; Kinase; Isoprene biosynthesis; ATP-binding;
 KW Complete proteome.
 FT NP BIND 88
 SO SEQUENCE 268 AA; 29788 MW; C905F8B988E0845D CRC64;
 Query Match
 Best Local Similarity 1.0%; Score 7; DB 1; Length 268;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 586 VETRISG 592
 |||||
 DB 152 VETRISG 158


```
RESULT 25
TEHB HAEIN STANDARD; PRT; 286 AA.
AC P45134;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tellurite resistance protein tehB homolog.
GN TEHB OR H1275.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";
RL Science 269:496-512 (1995).
CC -!- FUNCTION: INVOLVED IN POTASSIUM TELLURITE RESISTANCE (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR: EMBL; U32807; AAC22923.1; -.
DR TIGR; H11275; -.
DR InterPro; IPR000051; SAM_bind.
DR InterPro; IPR004537; TehB.
DR TIGRFAMs; TIGR00477; tehB; 1.
KW Antibiotic resistance; Tellurium resistance; Complete proteome.
SQ SEQUENCE 286 AA; 33019 MW; 69D7647E97CC74FA CRC64;

Query Match 1.0%; Score 7; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 NSLYLSL 510
Db 134 NSLYLSL 140
|||||

RESULT 26
NADC_SINY3 STANDARD; PRT; 295 AA.
ID P74301;
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable nicotinate-nucleotide pyrophosphorylase [carboxylating]
DE (EC 2.4.2.19) (Quinolinate phosphoribosyltransferase
DE [decarboxylating]) (QAPrase).
GN NADC OR SLR0936.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.

Query Match 1.0%; Score 7; DB 1; Length 286;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 504 NSLYLSL 510
Db 134 NSLYLSL 140
|||||

RESULT 27
FEPB_ECOLI STANDARD; PRT; 318 AA.
ID FEPB_ECOLI STANDARD; PRT; 318 AA.
AC P14609;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferrienterobactin-binding periplasmic protein precursor.
GN FEPB OR B0592.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90008779; PubMed=2529253;
RA Elkins M.F., Eazhart C.F.;
RT "Nucleotide sequence and regulation of the Escherichia coli gene for
RT ferrienterobactin transport protein FepB.";
RL J. Bacteriol. 171:5443-5451 (1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
```



```

RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RA  Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,
RA  Federspiel N., Hyman R., Katman S., Kemp C., Kuri O., Lew H., Lin D.,
RA  Nameth A., Oetner P., Roberts D., Schramm S., Davis R.W.;
RL  Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN  [4]
RP  SEQUENCE OF 1-57 FROM N.A.
RX  MEDLINE=90230305; PubMed=2139473;
RA  Brickman T.J., Osenberger B.A., McIntosh M.A.;
RT  "Regulation of divergent transcription from the iron-responsive fepB-
RT  entc promoter-operator regions in Escherichia coli.";
RL  J. Mol. Biol. 212:669-682(1990).
RN  [5]
RP  CHARACTERIZATION.
RX  MEDLINE=96004464; PubMed=7551033;
RA  Stephens D.L., Choe M.D., Earhart C.F.;
RT  "Escherichia coli periplasmic protein fepB binds ferriterobactin.";
RL  Microbiology 141:1647-1654(1995).
CC  -1- FUNCTION: BINDS FERRITEROBACTIN. PART OF THE BINDING-PROTEIN-
CC  -1- SUBCELLULAR LOCATION: Periplasmic.
CC  -1- DEPENDENT TRANSPORT SYSTEM FOR UPTAKE OF FERRITEROBACTIN.
CC  -1- INDUCTION: CONTROLLED IN PART BY THE AMOUNT OF AVAILABLE IRON.
CC  -1- SIMILARITY: BELONGS TO THE BACTERIAL EXTRACELLULAR SOLUTE-BINDING
CC  PROTEIN FAMILY 8.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M29730; AAA83853.1; -
DR  EMBL; AB000164; AAC73693.1; -
DR  EMBL; U82598; AAB40791.1; -
DR  EMBL; X53274; CA337370.1; -
DR  PIR; JY0045; JY0045.
DR  EcoGene; EG10294; fepB.
DR  InterPro; IPR002491; Peripla_BP_2.1.
DR  Pfam; PF01497; Peripla_BP_2.1.
KW  Transport; Iron transport; Signal; Periplasmic; Complete proteome.
FT  SIGNAL 1 26
FT  CHAIN 27 318 FERRITEROBACTIN-BINDING PERIPLASMIC
FT  PROTEIN.
SQ  SEQUENCE 318 AA; 34283 MW; 807676D03A508079 CRC64;
QY  Query Match 1.0%; Score 7; DB 1; Length 318;
   Best Local Similarity 100.0%; Pred. No. 53;
   Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  629 HSNALMT 635
   |||||
Db  204 HSNALMT 210

```

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Pm70;
RX  MEDLINE=21145866; PubMed=11248100;
RA  May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT  "Complete genomic sequence of Pasteurella multocida Pm70.";
RL  Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC  -1- CATALYTIC ACTIVITY: Anthranilate + phosphoribosylpyrophosphate =
CC  N-5'-phosphoribosyl-anthranilate + diphosphate.
CC  -1- PATHWAY: Tryptophan biosynthesis; second step.
CC  -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC  FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AB006093; AAK02665.1; -
DR  InterPro; IPR000312; Glycose_transf_3.
DR  Pfam; PF00591; Glycose_transf_3.1.
DR  Pfam; PF02885; Glycose_transf_3n.1.
DR  ProDom; PD001864; Glycose_transf_3.1.
DR  ProDom; PD005916; Thymid phosphils; 1.
DR  TIGRFAMs; TIGR01245; trpD; 1.
KW  Tryptophan biosynthesis; Transferase; Glycosyltransferase;
KW  Complete proteome.
SQ  SEQUENCE 333 AA; 35880 MW; F42760CD105F6C5A CRC64;
QY  Query Match 1.0%; Score 7; DB 1; Length 333;
   Best Local Similarity 100.0%; Pred. No. 55;
   Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY  589 EISGAVT 595
   |||||
Db  54 EISGAVT 60

```

```

RESULT 29
ID  OTCC_HAEIN STANDARD; PRT; 334 AA.
AC  P44770;
DT  01-NOV-1995 (Rel. 32, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Ornithine carbamoyltransferase, catabolic (EC 2.1.3.3) (OTCase).
GN  ARCB OR HI0596.
OS  Haemophilus influenzae.
OC  Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC  Haemophilus.
OX  NCBI_TaxID=727;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Rd / KW20 / ATCC 51907;
RX  MEDLINE=95550630; PubMed=7542800;
RA  Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA  Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA  McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA  Scott J.D., Shiley R., Liu L.-I., Glodet A., Kelley J.M.,
RA  Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA  Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA  Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA  Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA  Venter J.C.;
RT  "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT  Rd.";
RL  Science 269:496-512(1995).
CC  -1- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-ornithine = phosphate
CC  + L-citrulline.
CC  -1- PATHWAY: Arginine degradation via arginine deiminase; second step.

```


CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE ATCASES/OTCASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32741; AAC22253.1; -.
DR HSSP: P08308; 1ORT.
DR TIGR: H10596; -.
DR InterPro: IPR002029; Asp/Orn Cotransf.
DR InterPro: IPR002292; Orn carboxylase.
DR Pfam: PF00185; OTCace; 1.
DR Pfam: PF02729; OTCace_N; 1.
DR PRINTS: PR00100; AOTCASE.
DR TIGRFAMs: TIGR00658; orni_carb_tr; 1.
DR TRANSFAM: TIGR00658; orni_carb_tr; 1.
KW Transferase: Arginine metabolism; Complete proteome.
SQ SEQUENCE 334 AA; 37647 MW; B44B9ED2C1B857F5 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 334;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 637 YQVTPPEL 643
DB 254 YQVTPPEL 260

RESULT 30
YAAA_SCHPO
ID YAAA_SCHPO STANDARD; PRT; 344 AA.
AC Q09801;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C2267.10 in chromosome I.
GN SPAC2267.10.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harzill D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitz E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares S., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lenhardt R., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe I., McCombie W.R., Paulsen I., Potashkin J.,
RA Spakovski G.V., Uesery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SIMILARITY: TO YEAST PIP1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z54328; CAA91134.1; -.
DR Hypothetical protein.
SQ SEQUENCE 344 AA; 37290 MW; 571D391D0F88F27D CRC64;

Query Match 1.0%; Score 7; DB 1; Length 344;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 655 ITSSAGM 661
DB 272 ITSSAGM 278

RESULT 31
PHOE_CITFR
ID PHOE_CITFR STANDARD; PRT; 351 AA.
AC Q01605;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer membrane pore protein E precursor.
GN PHOE.
OS Citrobacter freundii.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Citrobacter.
OX NCBI_TaxID=546;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93146376; PubMed=1337052;
RA Spierings G., Ockhuijsen C., Hofstra H., Tommassen J.,
RT "Characterization of the Citrobacter freundii phoE gene and
RT development of C. freundii-specific oligonucleotides."
RL FEMS Microbiol. Lett. 78:199-204(1992).
CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE
CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS
CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,
CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED
CC SOLUTES.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO THE OMP/PHOE FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X68021; CAA48162.1; -.
DR PIR: S25520; S25520.
DR HSSP: P02932; IPHO.
DR InterPro: IPR003229; OMP_2.
DR InterPro: IPR001702; Porin gram-v.
DR Pfam: PF00267; Gram-v porins; 1.
DR PRINTS: PR00182; ECOLNEIPORIN.
DR PROSITE: PS00576; GRAM_NEG_PORIN; 1.

KM Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 351
 SQ SEQUENCE 351 AA; 38807 MW; 4E9C14A070381B5A CRC64;

Query Match 1.0%; Score 7; DB 1; Length 351;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 328 AFAGSKL 334
 DB 98 AFAGSKL 104

RESULT 32
 ID MURB TREPA STANDARD; PRT; 354 AA.
 AC 083128;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE UDP-N-acetylglucosaminylglucosamine reductase (EC 1.1.1.158) (UDP-N-acetylmuramate dehydrogenase).
 GN MURB OR TP0090.
 OS Treponema pallidum.
 OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
 OX NCBI_TaxId=160;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nichols;
 RX MEDLINE=98332770; PubMed=9665876;
 RA Frazer C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G., Dodson R., Gwin M., Hickey E.K., Clayton R., Ketchum K.A., Sodergren E., Hardham J.M., McLeod M.P., Salberg S., Peterson J., Khalak H., Richardson D., Howell J.K., Chidambaram M., Uterback T., McDonald L., Artich P., Bowman C., Cotton M.D., Fujii C., Garland S., Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O., Venter J.C.;
 RA "Complete genome sequence of Treponema pallidum, the syphilis vector spirochete."
 RT Science 281:375-388(1998).
 RL CC
 CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
 CC -1- COFACTOR: FAD (BY SIMILARITY).
 CC -1- PATHWAY: peptidoglycan biosynthesis.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (probable).
 CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AE001193; AAC65086.1; -.
 CC TIGR; TP0090;
 DR InterPro; IPR003170; MurB.
 DR InterPro; IPR001575; Oxid_FAD_bind.
 DR Pfam; PF01565; FAD binding_4; 1.
 DR Pfam; PF02873; MurB_C; 1.
 KW Peptidoglycan synthesis; Cell wall; Cell division; Oxidoreductase;
 KW NADP; Flavoprotein; FAD; Complete proteome.
 SQ SEQUENCE 354 AA; 38595 MW; 809E5B606F5BFF2 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 354;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 461 LSLIGGG 467
 I|||||

DB 56 LSLIGGG 62

RESULT 33
 ID V363 ASFB7 STANDARD; PRT; 363 AA.
 AC P23164;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE D'363 protein.
 GN DP363R.
 OS African swine fever virus (strain BA71V) (ASFV).
 OC Viruses; dsDNA viruses, no RNA stage; Asfarviridae; Asfivirus.
 OX NCBI_TaxId=10498;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90219205; PubMed=2325203;
 RA Gonzalez A., Calvo V., Almazan F., Almeral J.M., Ramirez J.C., de la Vega I., Blasco R., Vinnela E.;
 RT "Multigene families in African swine fever virus: family 360."
 RL J. Virol. 64:2073-2081(1990).
 RN [2]
 RP COMPLETE GENOME.
 RA Yanez R.J., Rodriguez J.M., Nogal M.L., Yuste L., Enriquez C., Rodriguez J.F., Vinnela E.;
 RT "Analysis of the complete nucleotide sequence of African swine fever virus."
 RL Virology 208:249-278(1995).
 CC
 CC -1- FUNCTION: NOT KNOWN.
 CC -1- SIMILARITY: BELONGS TO THE MULTIGENE FAMILY 360.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M57545; AAA42678.1; -.
 CC EMBL; U18466; AAA65382.1; -.
 DR PIR; F43680; F43680.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR002595; ASFV_360.
 DR Pfam; PR00023; ank; 1.
 DR Pfam; PF01671; ASFV_360; 1.
 DR PRODOM; PD003462; ASFV_360; 1.
 DR SMART; SM00248; ANK; 1.
 KW Multigene family.
 KW CONFICT 241 M -> T (IN REF. 1).
 FT SEQUENCE 363 AA; 42353 MW; 4841A8165447FE3 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 363;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 143 NLFAPDR 149
 DB 349 NLFAPDR 355

RESULT 34
 ID IPAC SHIDY STANDARD; PRT; 382 AA.
 AC Q03946;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 42 kDa membrane antigen precursor.
 GN IPAC.
 OS Shigella dysenteriae.
 OG Plasmid invasion.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=622;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CG097;
 RX MEDLINE=921114800; PubMed=1766387;
 RA Yao R., Palchaudhuri S.;
 RT "Nucleotide sequence of the ipaBCD structural genes of Shigella
 dysenteriae.";
 RL Mol. Microbiol. 5:2217-2221(1991).
 CC -!- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
 CC EPITHELIAL CELLS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X60777; CAA43191.1; -.
 DR PIR; S15578; S15578.
 KW Antigen; Plasmid; Virulence; Membrane; Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT CHAIN 20 382 42 KDA MEMBRANE ANTIGEN.
 SQ SEQUENCE 382 AA; 41098 MW; A22E1D6399EC90BF CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LAALSSS 24
 DB 170 LAALSSS 176
 RESULT 35
 IPAC_SHIFL STANDARD; PRT; 382 AA.
 AC P18012;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 42 kDa membrane antigen precursor.
 GN IPAC.
 OS Shigella flexneri.
 OG Plasmid 210 kb invasion pWR100, and plasmid 230 kb pMYSH6000.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Shigella.
 OX NCBI_TaxID=623;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=89057927; PubMed=3057506;
 RA Venkatesan M.M., Buysse J.M., Kopecko D.J.;
 RT "Characterization of invasion plasmid antigen genes (ipaBCD) from
 Shigella flexneri.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9317-9321(1988).
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M90T / Serotype 5;
 RX MEDLINE=89200844; PubMed=3071655;
 RA Baudry B., Kaczorek M., Sansonetti P.J.;
 RT "Nucleotide sequence of the invasion plasmid antigen B and C genes
 (ipaB and ipaC) of Shigella flexneri.";
 RL Microb. Pathog. 4:345-357(1988).
 RN (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype 2A; PLASMID=230 kb pMYSH6000;
 RX MEDLINE=90014179; PubMed=2552264;
 RA Sasakawa C., Adler B., Tobe T., Okada N., Nagai S., Komatsu K.,

RA Yoshikawa M.;
 RT "Functional organization and nucleotide sequence of virulence
 Region-2 on the large virulence plasmid in Shigella flexneri 2a.";
 RL Mol. Microbiol. 3:1191-1201(1989).
 RN (4)
 RP PARTIAL SEQUENCE OF 20-64 AND 318-335.
 RX MEDLINE=89307550; PubMed=2663721;
 RA Sankaran K., Ramachandran V., Subrahmanyam Y.V.B.K., Rajarathnam S.,
 RA Sankaran S., Roy R.K.;
 RT "Congo red-mediated regulation of levels of Shigella flexneri 2a
 membrane proteins.";
 RL Infect. Immun. 57:2364-2371(1989).
 CC -!- FUNCTION: ASSOCIATED WITH THE ENTRY OF THE BACTERIA INTO COLONIC
 CC EPITHELIAL CELLS.
 CC -----
 CC -!- MISCELLANEOUS: SYNTHESIS OF THIS IMMUNOGEN IS RERESSED AT 30
 CC DEGREES CELSIUS AND RESTORED AT 37 DEGREES CELSIUS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J04117; AAA26523.1; -.
 DR EMBL; M34849; AAA98425.1; ALT INIT.
 DR EMBL; X15319; CAA33382.1; -.
 DR PIR; A60112; A60112.
 KW Antigen; Plasmid; Virulence; Membrane; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 382 42 KDA MEMBRANE ANTIGEN.
 FT CONFLICT 30 30 T -> I (IN REF. 3).
 FT CONFLICT 372 372 A -> T (IN REF. 3).
 SQ SEQUENCE 382 AA; 41038 MW; ED6481B161E58EB1 CRC64;
 Query Match 1.0%; Score 7; DB 1; Length 382;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 18 LAALSSS 24
 DB 170 LAALSSS 176
 RESULT 36
 BN51_HUMAN STANDARD; PRT; 395 AA.
 ID BN51_HUMAN
 AC P05423;
 DT 01-NOV-1988 (Rel. 09, Created)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE BN51 protein.
 GN BN51T OR BN51.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=88065472; PubMed=3693386;
 RA Ittmann M., Greco A., Basilico C.;
 RT "Isolation of the human gene that complements a temperature-sensitive
 cell cycle mutation in BHK cells.";
 RL Mol. Cell. Biol. 7:3386-3393(1987).
 RN (2)
 RP SEQUENCE OF 1-12 FROM N.A.
 RA Ittmann M.;
 RT "Promoter structure and cell cycle control of the BN51 cell cycle
 gene, which encodes a subunit of RNA polymerase III.";
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPLEMENTS A TEMPERATURE-SENSITIVE CELL CYCLE MUTATION
 CC IN BHK CELLS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M17754; AAA51838.1; -
CC EMBL: L15301; AAA72377.1; -
CC PIR: A43700; A43700.
CC GeneW; HGNC:1080; BMS1T.
CC MIM: 187280; -
CC FT DOMAIN 62 98 HYDROPHILIC.
CC FT DOMAIN 138 149 HYDROPHILIC.
CC FT DOMAIN 215 230 HYDROPHILIC.
CC FT DOMAIN 289 306 HYDROPHILIC.
CC FT SIMILAR 289 395 TO THE C-TERMINUS OF YEAST RPC4.
SQ SEQUENCE 395 AA; 43978 MW; 897DEC72571844D CRC64;

Query Match 1.0%; Score 7; DB 1; Length 395;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ARKTPGL 103
DB 236 ARKTPGL 242

RESULT 37
VGLX HSVB STANDARD; PRT; 411 AA.
ID P28967;
AC 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein G precursor.
GN 70.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCB1_Taxid=31520;
RN [1]
RP MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -1- SIMILARITY: TO PRV GLYCOPROTEIN X.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M86664; AAB02505.1; -
CC EMBL: G36802; VGBE2.
CC InterPro; IPR003363; Herpes_gg.
DR Pfam; PF02400; Herpes_gg; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 411 GLYCOPROTEIN G.
FT TRANSMEM 363 379 POTENTIAL.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 392 392 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQ SEQUENCE 411 AA; 45269 MW; CC6349C38D0E12E2 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TVLAALS 22
DB 3 TVLAALS 9

RESULT 38
VGLX HSVB STANDARD; PRT; 411 AA.
ID P32514;
AC 01-OCT-1993 (Rel. 27, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycoprotein G precursor.
GN US3.
OS Equine herpesvirus type 1 (strain Kentucky A) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCB1_Taxid=10329;
RN [1]
RP MEDLINE=92263758; PubMed=1316673;
RA Colle C.F., Flowers C.C., O'Callaghan D.J.;
RT "Open reading frames encoding a protein kinase, homolog of
RT glycoprotein gX of pseudorabies virus, and a novel glycoprotein map
RT within the unique short segment of equine herpesvirus type 1."
RL Virology 188:545-557(1992).
RN [2]
RP REVISIONS TO C-TERMINUS.
RA Kinkou M., Fukushi H., Matsumura T., Kim S.K., O'Callaghan D.J.;
RL Submitted (SEP-2001) to the EMBL/Genbank/DBS databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M87497; AAA46071.2; -
CC PIR: B42538; VGBEKA.
CC InterPro; IPR003363; Herpes_gg.
DR Pfam; PF02400; Herpes_gg; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 411 GLYCOPROTEIN GX.
FT TRANSMEM 364 384 POTENTIAL.
FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 317 317 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 411 AA; 45327 MW; CESDB8348CACC1144 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 411;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TVLAALS 22
DB 3 TVLAALS 9

RESULT 39
GLUP BRUAB STANDARD; PRT; 412 AA.
ID GLUP BRUAB
AC Q44623;


```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Glucose/galactose transporter.
GN GLUP.
OS Brucella abortus.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=235;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2308;
RX MEDLINE=97311994; PubMed=9168605;
RA Essenberg R.C., Candler C., Nida S.K.;
RT "Brucella abortus strain 2308 putative glucose and galactose
RT transporter gene: cloning and characterization.";
RL Microbiology 143:1549-1555(1997).
CC -!- FUNCTION: INTAKE OF GLUCOSE AND GALACTOSE (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Probable).
CC -!- SIMILARITY: BELONGS TO THE FHS (TC 2.1.7) FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U43785; AAB58958.1; -.
CC TIGRFAMs; TIGR01272; Glup; 1.
KW Sugar transport; Transmembrane; Inner membrane.
FT TRANSMEM 21 41 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 158 178 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 239 259 POTENTIAL.
FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 331 351 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 388 408 POTENTIAL.
SQ SEQUENCE 412 AA; 43859 MW; 716D3AAB9E197E5F CRC64;

Query Match 1.0%; Score 7; DB 1; Length 412;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 364 KALAFDA 370
DB 304 KALAFDA 310

RESULT 40
Y118 TREPA
ID Y118 TREPA STANDARD; PRT; 421 AA.
AC O83155;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0118.
GN TP0118.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

```

```

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khaliq H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE001197; AAC65114.1; -.
CC TIGR; TP0118; -.
KW Hypothetical protein; Coiled coil; Complete proteome.
FT DOMAIN 126 182 COILED COIL (POTENTIAL).
FT DOMAIN 328 397 COILED COIL (POTENTIAL).
SQ SEQUENCE 421 AA; 46702 MW; 8C7C72BD0C515B6A CRC64;

Query Match 1.0%; Score 7; DB 1; Length 421;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 576 AALGKRV 582
DB 82 AALGKRV 88

RESULT 41
FOLC HAEIN STANDARD; PRT; 437 AA.
ID FOLC HAEIN STANDARD; PRT; 437 AA.
AC P43775;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Folylpolyglutamate synthase (EC 6.3.2.17) (Folylpoly-gamma-glutamate
DE synthetase) (FPGS).
GN FOLC OR H1261.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Uddelman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: CONVERSION OF FOLATES TO POLYGLUTAMATE DERIVATIVES
CC (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + [tetrahydrofolyl-[Glu]](N) + L-glutamate
CC = ADP + phosphate + [tetrahydrofolyl-[Glu]](N+1).
CC -!- PATHWAY: BACTERIA REQUIRE FOLATE FOR THE BIOSYNTHESIS OF GLYCINE,
CC METHIONINE, FORMYL-MET-TRNA, THYMIDYLATES, PURINES, AND
CC PANTOTHENATE.
CC -!- SIMILARITY: BELONGS TO THE FOLYLPOLYGLUTAMATE SYNTHASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```



```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32806; AAC22914.1; -.
DR HSSP; P15925; 1FGS.
DR TIGR; H11261; -.
DR InterPro; IPR001645; Fpolylgl synthase.
DR InterPro; IPR000713; Mur_ligase.
DR Pfam; PF01225; Mur_ligase; 1.
DR PROSITE; PS01011; FOLYLPOLYGLU SYNT 1; 1.
DR PROSITE; PS01012; FOLYLPOLYGLU SYNT 2; 1.
DR LIGASE; One-carbon metabolism; ATP-binding; Complete proteome.
KW NP BIND 55 ATP (BY SIMILARITY).
FT SEQUENCE 437 AA; 48826 MW; 04A493395E76CD43 CRC64;
SQ
Query Match 1.0%; Score 7; DB 1; Length 437;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVLA 20
Db 256 AATVLA 262

RESULT 42
UVSH_EMENT STANDARD; PRT; 443 AA.
ID UVSH_EMENT
AC Q02398; Q00178;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Postreplication repair protein uvsh/nuva.
GN UVSH OR NUVA.
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxId=5072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FGSC 4;
RX MEDLINE=95379762; PubMed=7651340;
RA Yoon J.H., Lee B.J., Kang H.S.;
RT "The Aspergillus uvsh gene encodes a product homologous to yeast RAD18
RT and Neurospora UVS-2."
RL Mol. Genet. 248:174-181(1995).
RN [2]
RP SEQUENCE OF 1-356 FROM N.A.
RC STRAIN=L20;
RX MEDLINE=97022065; PubMed=8868425;
RA Iwawake L.A., Cotton C.M., Jones G.W., Tomasett A.B., Strike P.;
RT "nuva", an Aspergillus nidulans gene involved in DNA repair and
RT recombination, is a homologue of Saccharomyces cerevisiae RAD18 and
RT Neurospora crassa uvs-2."
RL Microbiology 142:505-515(1996).
CC -!- FUNCTION: Involved in postreplication repair of UV-damaged DNA.
CC Postreplication repair functions in gap-filling of a daughter
CC strand on replication of damaged DNA. Has ssDNA binding activity
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL; S79392; AAB35098.1; -.
DR EMBL; Z49875; CAA90033.1; -.
DR HSSP; P15919; 1RMD.
DR InterPro; IPR004580; RAD18.
DR InterPro; IPR003034; SAP.
DR InterPro; IPR01841; Znf_ring.
DR Pfam; PF00097; zf-C3HC4; 1.
DR Pfam; PF02037; SAP; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00513; SAP; 1.
DR TIGRfam; TIGR00599; rad18; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
DR Zinc-finger; DNA-binding; DNA damage; DNA repair; Nuclear protein.
FT ZN FING 30 68
FT DOMAIN 236 270
FT CONFLICT 209 210 FG -> YR (IN REF. 2).
SQ SEQUENCE 443 AA; 48906 MW; DD3327065D3511B2 CRC64;

QY 424 LIANARK 430
Db 334 LIANARK 340

RESULT 43
SR54_AQUAE STANDARD; PRT; 454 AA.
ID SR54_AQUAE
AC O67615;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Signal recognition particle protein (fifty-four homolog).
GN PFH OR AQ_1720.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxId=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
RA Felman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT EXPORT OF EXTRA-CYTOPLASMIC
CC PROTEINS. BINDS TO THE SIGNAL SEQUENCE WHEN IT EMERGES FROM THE
CC RIBOSOMES (BY SIMILARITY).
CC -!- SUBUNIT: SIGNAL RECOGNITION PARTICLE CONSISTS OF A 4.5 RNA
CC MOLECULE AND PROTEIN PFH (BY SIMILARITY).
CC -!- DOMAIN: THE PROTEIN HAS A TWO DOMAIN STRUCTURE. THE G-DOMAIN
CC BINDS GTP. THE M-DOMAIN BINDS THE RNA AND ALSO BINDS THE SIGNAL
CC SEQUENCE.
CC -!- SIMILARITY: BELONGS TO THE SRP FAMILY OF GTP-BINDING PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000753; AAC07579.1; -.
DR HSSP; O07347; 1FRH.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR000897; SRP54.

```



```

DR InterPro; IPR004125; SRP54_SPB.
DR InterPro; IPR004780; SRP_sub.
DR Pfam; PF00448; SRP54; 1.
DR Pfam; PF02881; SRP54; 1.
DR Pfam; PF02978; SRP_SPB; 1.
DR ProDom; PD000819; SRP54; 1.
DR SMART; SM00382; RAA; 1.
DR TIGRFAMs; TIGR00959; 3a0501s01; 1.
DR PROSITE; PS00300; SRP54; 1.
KW Signal recognition particle; GTP-binding; RNA-binding;
KW Complete proteome.
FT DOMAIN 1 289 G-DOMAIN.
FT DOMAIN 290 454 M-DOMAIN.
FT NP_BIND 102 109 GTP (BY SIMILARITY).
FT NP_BIND 184 188 GTP (BY SIMILARITY).
FT NP_BIND 242 245 GTP (BY SIMILARITY).
SQ SEQUENCE 454 AA; 50911 MW; 9F13FA8C883A0558 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 454;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
DB 248 ARGGLAL 254

RESULT 44
ID EF1A_DICDI STANDARD; PRT; 456 AA.
AC P18624;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor 1-alpha (EF-1-alpha) [50 kDa actin-binding protein]
DE (ABP-50).
GN EF1A.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
RX MEDLINE=91015340; PubMed=2215665;
RA Yang F., Demma M., Warren V., Dharmawardane S., Condeelis J.;
RT "Identification of an actin-binding protein from Dictyostelium as
RT elongation factor 1a.";
RL Nature 347:494-496(1990).
CC -!- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-tRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS. IT IS ALSO AN ABUNDANT ACTIN FILAMENT BUNDLING
CC PROTEIN.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; X55972; CAA39442.1; -.
DR EMBL; X55973; CAA39443.1; -.
DR PIR; S11665; S11665.
DR HSP; P07157; 1AIP.
DR DictyDb; DD01027; efaA.
DR InterPro; IPR004539; EF1_alpha.
DR InterPro; IPR004160; EFTU_Cterm.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.

```

```

DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03143; GTP_EFTU_D3; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR TIGRFAMs; TIGR00483; EF-1_alpha; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding; Actin-binding.
FT NP_BIND 17 24 GTP (BY SIMILARITY).
FT NP_BIND 94 98 GTP (BY SIMILARITY).
FT NP_BIND 156 159 GTP (BY SIMILARITY).
SQ SEQUENCE 456 AA; 50066 MW; BCFE9E1B70C3D8A8 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 QIHAGYS 605
DB 343 QIHAGYS 349

RESULT 45
TBB1_PORPU STANDARD; PRT; 457 AA.
AC P50259;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tubulin beta-1 chain.
GN TUBB1.
OS Porphyra purpurea.
OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.
OX NCBI_TaxID=2787;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Avonport;
RA Mackay R.M., Gallant J.W.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -!- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -!- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; Z67991; CAA91939.1; -.
DR InterPro; IPR002453; Beta_tubulin.
DR InterPro; IPR000217; Tubulin.
DR InterPro; IPR003008; Tubulin_Fts2.
DR Pfam; PF00091; tubulin; 1.
DR PRINTS; PR01161; TUBULIN.
DR PROSITE; PS00227; TUBULIN; 1.
DR PROSITE; PS00228; TUBULIN_B_AUTOREG; 1.
KW Microtubules; GTP-binding; Multigene family.
FT NP_BIND 140 146 GTP (POTENTIAL).
SQ SEQUENCE 457 AA; 50047 MW; 0F2E5BFA7E077DAA CRC64;

Query Match 1.0%; Score 7; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 652 MSGITSS 658
DB 233 MSGITSS 239

```


Search completed: December 18, 2002, 06:57:25
Job time : 30.2194 secs

270

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:57:00 ; Search time 12.7742 Seconds
(without alignments)
1662.984 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 722

Sequence: 1 MCGFMSVFRINMTATVLA.....NTEPIGSERTWTANLRYSF 722

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep:*

2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep:*

3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep:*

4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep:*

5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep:*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.0	17	2	US-08-656-177A-3
2	7	1.0	17	4	US-09-256-797-3
3	7	1.0	75	4	US-08-817-811-97
4	7	1.0	76	4	US-08-817-811-21
5	7	1.0	78	4	US-08-817-811-20
6	7	1.0	78	4	US-08-817-811-96
7	7	1.0	79	4	US-08-817-811-19
8	7	1.0	79	4	US-08-817-811-95
9	7	1.0	127	1	US-08-392-828C-39
10	7	1.0	127	1	US-09-330-945-39
11	7	1.0	248	6	5182362-8
12	7	1.0	278	1	US-07-941-651-2
13	7	1.0	278	1	US-08-279-996-2
14	7	1.0	324	1	US-08-746-797-2
15	7	1.0	324	1	US-08-927-387-2
16	7	1.0	324	2	US-08-918-058-2
17	7	1.0	352	3	US-08-921-887-52
18	7	1.0	411	2	US-08-338-530A-3
19	7	1.0	411	2	US-09-267-384-3
20	7	1.0	461	4	US-09-457-046B-64
21	7	1.0	491	2	US-08-468-812-8
22	7	1.0	491	2	US-08-590-563-8
23	7	1.0	554	4	US-09-319-989-6
24	7	1.0	557	4	US-09-134-001C-5569
25	7	1.0	558	2	US-08-656-177A-2
26	7	1.0	558	4	US-09-256-797-2
27	7	1.0	661	4	US-09-564-805-240

28	7	1.0	714	4	US-09-347-878-16	Sequence 16, Appl
29	7	1.0	971	4	US-09-405-728-2	Sequence 2, Appl
30	7	1.0	1239	1	US-08-026-138B-3	Sequence 3, Appl
31	7	1.0	1385	2	US-08-687-399-7	Sequence 7, Appl
32	7	1.0	1544	4	US-09-413-814-46	Sequence 46, Appl
33	7	1.0	1618	4	US-08-462-467B-4	Sequence 4, Appl
34	7	1.0	2887	4	US-08-462-467B-2	Sequence 8, Appl
35	7	1.0	2887	4	US-08-462-467B-2	Sequence 8, Appl
36	6	0.8	10	3	US-08-159-339A-836	Sequence 836, App
37	6	0.8	10	3	US-08-159-339A-891	Sequence 891, App
38	6	0.8	17	1	US-08-191-866D-64	Sequence 64, Appl
39	6	0.8	17	2	US-08-185-949B-64	Sequence 64, Appl
40	6	0.8	17	5	PCT-US93-07306-45	Sequence 45, Appl
41	6	0.8	19	2	US-08-520-933-2	Sequence 2, Appl
42	6	0.8	19	4	US-09-285-040-3	Sequence 2, Appl
43	6	0.8	23	3	US-08-256-747C-59	Sequence 59, Appl
44	6	0.8	23	4	US-08-834-130A-59	Sequence 59, Appl
45	6	0.8	24	2	US-08-726-306A-118	Sequence 118, App
46	6	0.8	25	2	US-08-139-609-5	Sequence 5, Appl
47	6	0.8	26	1	US-08-394-326-12	Sequence 12, Appl
48	6	0.8	26	3	US-09-082-306-12	Sequence 12, Appl
49	6	0.8	28	2	US-09-010-928B-15	Sequence 15, Appl
50	6	0.8	38	1	US-08-176-500-88	Sequence 88, Appl
51	6	0.8	38	1	US-08-471-052A-88	Sequence 88, Appl
52	6	0.8	38	1	US-08-189-531-88	Sequence 88, Appl
53	6	0.8	38	2	US-08-471-939-88	Sequence 88, Appl
54	6	0.8	38	2	US-08-471-800-88	Sequence 88, Appl
55	6	0.8	38	2	US-08-471-068-88	Sequence 88, Appl
56	6	0.8	41	1	US-08-363-311-18	Sequence 18, Appl
57	6	0.8	41	1	US-08-363-311-20	Sequence 20, Appl
58	6	0.8	41	2	US-08-463-288A-18	Sequence 18, Appl
59	6	0.8	41	2	US-08-463-288A-20	Sequence 20, Appl
60	6	0.8	41	2	US-08-470-445A-18	Sequence 18, Appl
61	6	0.8	41	2	US-08-470-445A-20	Sequence 20, Appl
62	6	0.8	41	2	US-08-462-679-18	Sequence 18, Appl
63	6	0.8	41	2	US-08-462-679-20	Sequence 20, Appl
64	6	0.8	41	2	US-08-466-210A-18	Sequence 18, Appl
65	6	0.8	41	2	US-08-466-210A-20	Sequence 20, Appl
66	6	0.8	41	2	US-08-467-147A-18	Sequence 18, Appl
67	6	0.8	41	2	US-08-467-147A-20	Sequence 20, Appl
68	6	0.8	41	2	US-08-469-014-18	Sequence 18, Appl
69	6	0.8	41	2	US-08-469-014-20	Sequence 20, Appl
70	6	0.8	41	4	US-09-346-290-18	Sequence 18, Appl
71	6	0.8	41	4	US-09-346-290-20	Sequence 20, Appl
72	6	0.8	41	5	PCT-US93-10506A-18	Sequence 18, Appl
73	6	0.8	41	5	PCT-US93-10506A-20	Sequence 20, Appl
74	6	0.8	41	5	PCT-US93-10506-18	Sequence 18, Appl
75	6	0.8	41	5	PCT-US93-10506-20	Sequence 20, Appl
76	6	0.8	42	1	US-08-363-311-19	Sequence 19, Appl
77	6	0.8	42	2	US-08-463-288A-19	Sequence 19, Appl
78	6	0.8	42	2	US-08-470-445A-19	Sequence 19, Appl
79	6	0.8	42	2	US-08-462-679-19	Sequence 19, Appl
80	6	0.8	42	2	US-08-466-210A-19	Sequence 19, Appl
81	6	0.8	42	2	US-08-467-147A-19	Sequence 19, Appl
82	6	0.8	42	2	US-08-469-014-19	Sequence 19, Appl
83	6	0.8	42	4	US-09-346-290-19	Sequence 19, Appl
84	6	0.8	42	5	PCT-US93-10506A-19	Sequence 19, Appl
85	6	0.8	42	5	PCT-US93-10506-19	Sequence 19, Appl
86	6	0.8	50	1	US-08-363-311-21	Sequence 21, Appl
87	6	0.8	50	2	US-08-463-288A-21	Sequence 21, Appl
88	6	0.8	50	2	US-08-470-445A-21	Sequence 21, Appl
89	6	0.8	50	2	US-08-462-679-21	Sequence 21, Appl
90	6	0.8	50	2	US-08-466-210A-21	Sequence 21, Appl
91	6	0.8	50	2	US-08-467-147A-21	Sequence 21, Appl
92	6	0.8	50	2	US-08-469-014-21	Sequence 21, Appl
93	6	0.8	50	4	US-09-004-406C-25	Sequence 25, Appl
94	6	0.8	50	4	US-09-346-290-21	Sequence 21, Appl
95	6	0.8	50	5	PCT-US93-10506A-21	Sequence 21, Appl
96	6	0.8	55	5	PCT-US93-10506-21	Sequence 21, Appl
97	6	0.8	55	1	US-08-394-326-7	Sequence 7, Appl
98	6	0.8	55	3	US-09-082-306-7	Sequence 7, Appl
99	6	0.8	63	4	US-09-183-861-52	Sequence 52, Appl
100	6	0.8	63	4	US-09-022-765-52	Sequence 52, Appl

101	6	0.8	75	2	US-08-633-770A-5	Sequence 5, Appli	174	6	0.8	203	1	US-08-444-083-3	Sequence 3, Appli
102	6	0.8	98	2	US-09-010-528B-24	Sequence 24, Appl	175	6	0.8	203	1	US-08-286-304-3	Sequence 3, Appli
103	6	0.8	99	2	US-09-047-125-24	Sequence 24, Appl	176	6	0.8	203	1	US-08-442-745-3	Sequence 3, Appli
104	6	0.8	99	3	US-07-736-335B-24	Sequence 24, Appl	177	6	0.8	203	1	US-08-443-129-3	Sequence 3, Appli
105	6	0.8	100	4	US-09-091-725-35	Sequence 35, Appl	178	6	0.8	203	1	US-08-443-953-3	Sequence 3, Appli
106	6	0.8	104	1	US-08-241-853-4	Sequence 4, Appli	179	6	0.8	203	1	US-08-443-130-3	Sequence 3, Appli
107	6	0.8	104	2	US-08-850-917-4	Sequence 4, Appli	180	6	0.8	203	3	US-09-106-182-3	Sequence 3, Appli
108	6	0.8	109	4	US-09-240-274-57	Sequence 57, Appl	181	6	0.8	203	3	US-08-898-911-3	Sequence 3, Appli
109	6	0.8	112	4	US-09-504-358-24	Sequence 24, Appl	182	6	0.8	203	5	PCT-US95-04467-3	Sequence 3, Appli
110	6	0.8	112	4	US-09-954-314-24	Sequence 24, Appl	183	6	0.8	206	4	US-09-562-737-55	Sequence 55, Appl
111	6	0.8	114	3	US-08-838-682-12	Sequence 12, Appl	184	6	0.8	206	4	US-09-562-737-55	Sequence 55, Appl
112	6	0.8	114	4	US-08-895-914-12	Sequence 12, Appl	185	6	0.8	206	4	US-09-562-737-56	Sequence 56, Appl
113	6	0.8	114	4	US-09-357-710A-12	Sequence 12, Appl	186	6	0.8	206	4	US-09-562-737-57	Sequence 57, Appl
114	6	0.8	126	4	US-08-483-533-31	Sequence 31, Appl	187	6	0.8	206	4	US-09-562-737-58	Sequence 58, Appl
115	6	0.8	126	4	US-09-283-471A-31	Sequence 31, Appl	188	6	0.8	208	2	US-08-531-525-20	Sequence 20, Appl
116	6	0.8	129	4	US-09-134-001C-4545	Sequence 4545, Ap	189	6	0.8	208	2	US-08-718-270A-20	Sequence 20, Appl
117	6	0.8	136	2	US-08-675-508-5	Sequence 5, Appli	190	6	0.8	209	3	US-08-464-052-4	Sequence 4, Appli
118	6	0.8	138	4	US-08-746-397-11	Sequence 11, Appl	191	6	0.8	209	4	US-08-461-002-4	Sequence 4, Appli
119	6	0.8	139	1	US-08-278-729A-6	Sequence 6, Appli	192	6	0.8	209	4	US-08-689-411-4	Sequence 4, Appli
120	6	0.8	139	1	US-08-155-343A-6	Sequence 6, Appli	193	6	0.8	210	2	US-08-948-616-5	Sequence 5, Appli
121	6	0.8	139	1	US-08-406-672-6	Sequence 6, Appli	194	6	0.8	210	2	US-09-193-510-5	Sequence 5, Appli
122	6	0.8	139	1	US-08-643-563A-6	Sequence 6, Appli	195	6	0.8	210	4	US-09-368-403-5	Sequence 5, Appli
123	6	0.8	139	1	US-08-643-763A-6	Sequence 6, Appli	196	6	0.8	210	4	US-09-198-119C-17	Sequence 17, Appl
124	6	0.8	139	1	US-08-462-623-6	Sequence 6, Appli	197	6	0.8	210	4	US-09-198-119C-47	Sequence 47, Appl
125	6	0.8	139	1	US-08-451-953A-6	Sequence 6, Appli	198	6	0.8	211	1	US-08-276-852-34	Sequence 34, Appl
126	6	0.8	139	2	US-08-445-468A-6	Sequence 6, Appli	199	6	0.8	211	1	US-08-133-011-16	Sequence 16, Appl
127	6	0.8	139	2	US-08-461-397A-6	Sequence 6, Appli	200	6	0.8	211	1	US-08-322-730A-16	Sequence 16, Appl
128	6	0.8	139	2	US-08-912-088-6	Sequence 6, Appli	201	6	0.8	211	1	US-08-387-874-16	Sequence 16, Appl
129	6	0.8	139	3	US-08-278-730A-6	Sequence 6, Appli	202	6	0.8	211	1	US-08-899-575-34	Sequence 34, Appl
130	6	0.8	139	3	US-08-445-467-6	Sequence 6, Appli	203	6	0.8	211	1	US-08-899-575-34	Sequence 34, Appl
131	6	0.8	139	3	US-08-480-515A-6	Sequence 6, Appli	204	6	0.8	211	2	US-08-383-619-16	Sequence 16, Appl
132	6	0.8	139	4	US-08-414-033A-6	Sequence 6, Appli	205	6	0.8	211	4	US-08-907-739-16	Sequence 16, Appl
133	6	0.8	139	4	US-08-271-556A-4	Sequence 4, Appli	206	6	0.8	211	4	US-09-729-597-16	Sequence 16, Appl
134	6	0.8	139	4	US-08-440-894A-6	Sequence 6, Appli	207	6	0.8	211	5	PCT-US93-08364-16	Sequence 16, Appl
135	6	0.8	139	4	US-09-170-936-6	Sequence 6, Appli	208	6	0.8	211	5	PCT-US95-08743-34	Sequence 34, Appl
136	6	0.8	139	4	US-08-461-113-6	Sequence 6, Appli	209	6	0.8	212	2	US-08-916-770-3	Sequence 3, Appli
137	6	0.8	139	5	PCT-US92-01968-6	Sequence 6, Appli	210	6	0.8	212	2	US-09-102-593-3	Sequence 3, Appli
138	6	0.8	139	5	PCT-US93-07190-6	Sequence 6, Appli	211	6	0.8	212	4	US-09-293-724-3	Sequence 3, Appli
139	6	0.8	139	5	PCT-US93-07231-6	Sequence 6, Appli	212	6	0.8	213	2	US-08-916-770-1	Sequence 1, Appli
140	6	0.8	139	5	PCT-US93-08742-6	Sequence 6, Appli	213	6	0.8	213	2	US-09-102-593-1	Sequence 1, Appli
141	6	0.8	139	5	PCT-US93-08808-6	Sequence 6, Appli	214	6	0.8	213	2	US-08-941-263-3	Sequence 3, Appli
142	6	0.8	139	5	PCT-US93-08885-6	Sequence 6, Appli	215	6	0.8	213	3	US-09-186-250-2	Sequence 2, Appli
143	6	0.8	147	3	US-08-346-329A-61	Sequence 61, Appl	216	6	0.8	213	3	US-09-186-250-6	Sequence 6, Appli
144	6	0.8	154	1	US-08-426-627-20	Sequence 20, Appl	217	6	0.8	213	3	US-09-227-178-3	Sequence 3, Appli
145	6	0.8	157	4	US-09-025-769B-372	Sequence 372, App	218	6	0.8	213	4	US-09-470-449-3	Sequence 3, Appli
146	6	0.8	157	4	US-09-025-769B-373	Sequence 373, App	219	6	0.8	213	4	US-09-517-347-2	Sequence 2, Appli
147	6	0.8	159	3	US-09-027-449-53	Sequence 53, Appl	220	6	0.8	213	4	US-09-517-347-6	Sequence 6, Appli
148	6	0.8	159	3	US-08-804-444A-53	Sequence 53, Appl	221	6	0.8	213	4	US-09-293-724-1	Sequence 1, Appli
149	6	0.8	159	4	US-09-026-985-53	Sequence 53, Appl	222	6	0.8	213	4	US-09-518-232A-2	Sequence 2, Appli
150	6	0.8	159	4	US-09-121-952A-53	Sequence 53, Appl	223	6	0.8	213	4	US-09-518-232A-6	Sequence 6, Appli
151	6	0.8	159	4	US-09-234-340A-53	Sequence 53, Appl	224	6	0.8	213	4	US-09-517-358A-2	Sequence 2, Appli
152	6	0.8	163	2	US-08-687-916-16	Sequence 16, Appl	225	6	0.8	213	4	US-09-517-358A-6	Sequence 6, Appli
153	6	0.8	163	4	US-09-138-614-16	Sequence 16, Appl	226	6	0.8	216	3	US-08-824-692-24	Sequence 24, Appl
154	6	0.8	163	4	US-09-025-769B-282	Sequence 282, App	227	6	0.8	219	1	US-08-152-019A-32	Sequence 32, Appl
155	6	0.8	163	4	US-08-858-207A-451	Sequence 451, App	228	6	0.8	222	4	US-09-227-357-161	Sequence 161, App
156	6	0.8	170	3	US-09-186-250-4	Sequence 4, Appli	229	6	0.8	224	2	US-08-272-255-16	Sequence 16, Appl
157	6	0.8	170	3	US-09-186-250-8	Sequence 8, Appli	230	6	0.8	224	5	PCT-US95-08565-16	Sequence 16, Appl
158	6	0.8	170	4	US-09-517-347-4	Sequence 4, Appli	231	6	0.8	227	4	US-08-871-572B-14	Sequence 14, Appl
159	6	0.8	170	4	US-09-517-347-8	Sequence 8, Appli	232	6	0.8	228	4	US-09-134-001C-4083	Sequence 4083, Ap
160	6	0.8	170	4	US-09-129-030-40	Sequence 40, Appl	233	6	0.8	232	2	US-08-738-413B-11	Sequence 11, Appl
161	6	0.8	170	4	US-09-518-232A-4	Sequence 4, Appli	234	6	0.8	232	2	US-08-829-876-103	Sequence 103, App
162	6	0.8	170	4	US-09-518-232A-8	Sequence 8, Appli	235	6	0.8	232	4	US-09-361-434-7	Sequence 7, Appli
163	6	0.8	170	4	US-09-517-358A-4	Sequence 4, Appli	236	6	0.8	232	4	US-09-234-874A-103	Sequence 103, App
164	6	0.8	170	4	US-09-517-358A-8	Sequence 8, Appli	237	6	0.8	232	4	US-09-635-025-7	Sequence 7, Appli
165	6	0.8	175	4	US-08-973-961-7	Sequence 7, Appli	238	6	0.8	233	2	US-08-829-876-101	Sequence 101, App
166	6	0.8	175	4	US-08-973-961-8	Sequence 8, Appli	239	6	0.8	233	2	US-08-829-876-105	Sequence 105, App
167	6	0.8	179	4	US-08-483-533-33	Sequence 33, Appl	240	6	0.8	233	4	US-08-438-745-15	Sequence 15, Appl
168	6	0.8	179	4	US-09-283-471A-33	Sequence 33, Appl	241	6	0.8	233	4	US-08-438-745-15	Sequence 15, Appl
169	6	0.8	182	1	US-08-611-107-2	Sequence 2, Appli	242	6	0.8	233	4	US-09-219-019-15	Sequence 15, Appl
170	6	0.8	182	2	US-08-422-560A-2	Sequence 2, Appli	243	6	0.8	233	4	US-09-219-019-17	Sequence 17, Appl
171	6	0.8	182	4	US-08-468-793-2	Sequence 2, Appli	244	6	0.8	233	4	US-09-234-874A-101	Sequence 101, App
172	6	0.8	200	4	US-09-134-001C-3099	Sequence 3099, Ap	245	6	0.8	233	4	US-09-234-874A-105	Sequence 105, App
173	6	0.8	203	1	US-08-233-609-3	Sequence 3, Appli	246	6	0.8	233	4	US-09-134-001C-3643	Sequence 3643, Ap

247	6	0.8	233	5	PCT-US94-05669A-15	Sequence 15, Appl	320	6	0.8	227	4	US-09-101-126-3	Sequence 3, Appl
248	6	0.8	233	5	PCT-US94-05669A-17	Sequence 17, Appl	321	6	0.8	227	4	US-09-367-528A-1	Sequence 1, Appl
249	6	0.8	233	5	US-08-438-745-13	Sequence 13, Appl	322	6	0.8	237	4	US-09-367-528A-3	Sequence 3, Appl
250	6	0.8	235	4	US-09-219-019-13	Sequence 13, Appl	323	6	0.8	227	4	US-09-367-528B-5	Sequence 5, Appl
251	6	0.8	235	5	PCT-US94-05669A-13	Sequence 13, Appl	324	6	0.8	229	4	US-08-353-476-68	Sequence 68, Appl
252	6	0.8	236	5	PCT-US95-04801-8	Sequence 8, Appl	325	6	0.8	303	4	US-09-134-001C-4369	Sequence 4369, Ap
253	6	0.8	239	4	US-08-896-933-27	Sequence 27, Appl	326	6	0.8	304	2	US-08-424-641B-4	Sequence 4, Appl
254	6	0.8	239	4	US-09-314-235-27	Sequence 27, Appl	327	6	0.8	304	2	US-08-820-980-4	Sequence 4, Appl
255	6	0.8	240	3	US-08-824-692-23	Sequence 23, Appl	328	6	0.8	304	4	US-08-826-439-4	Sequence 4, Appl
256	6	0.8	241	4	US-09-370-838-194	Sequence 194, App	329	6	0.8	304	4	US-08-913-159-8	Sequence 8, Appl
257	6	0.8	241	4	US-09-370-838-204	Sequence 204, App	330	6	0.8	305	2	US-08-946-528-1	Sequence 1, Appl
258	6	0.8	241	4	US-09-387-372-4	Sequence 4, Appl	331	6	0.8	310	4	US-09-632-847B-8	Sequence 8, Appl
259	6	0.8	243	4	US-09-413-814-20	Sequence 20, Appl	332	6	0.8	311	4	US-09-105-390-44	Sequence 44, Appl
260	6	0.8	245	4	US-08-483-533-42	Sequence 42, Appl	333	6	0.8	313	3	US-08-926-842B-62	Sequence 62, Appl
261	6	0.8	245	4	US-09-342-084-4	Sequence 4, Appl	334	6	0.8	316	4	US-09-241-750-2	Sequence 2, Appl
262	6	0.8	245	4	US-09-283-471A-42	Sequence 42, Appl	335	6	0.8	318	4	US-09-347-803-4	Sequence 4, Appl
263	6	0.8	251	1	US-08-209-747-8	Sequence 8, Appl	336	6	0.8	319	4	US-09-347-803-2	Sequence 2, Appl
264	6	0.8	251	1	US-08-488-298-8	Sequence 8, Appl	337	6	0.8	319	4	US-09-347-803-24	Sequence 24, Appl
265	6	0.8	252	1	US-08-271-354-14	Sequence 14, Appl	338	6	0.8	320	4	US-09-347-803-23	Sequence 23, Appl
266	6	0.8	252	1	US-08-565-861-14	Sequence 14, Appl	339	6	0.8	321	2	US-08-937-972-3	Sequence 3, Appl
267	6	0.8	252	4	US-08-483-533-43	Sequence 43, Appl	340	6	0.8	321	4	US-09-361-434-17	Sequence 17, Appl
268	6	0.8	252	4	US-09-283-471A-43	Sequence 43, Appl	341	6	0.8	321	4	US-09-361-434-22	Sequence 22, Appl
269	6	0.8	252	5	PCT-US94-07658-14	Sequence 14, Appl	342	6	0.8	321	4	US-09-635-025-17	Sequence 17, Appl
270	6	0.8	258	4	US-08-483-533-26	Sequence 26, Appl	343	6	0.8	321	4	US-09-635-025-22	Sequence 22, Appl
271	6	0.8	258	4	US-09-283-471A-26	Sequence 26, Appl	344	6	0.8	325	4	US-09-267-031-14	Sequence 14, Appl
272	6	0.8	258	4	US-09-430-564-6	Sequence 6, Appl	345	6	0.8	327	4	US-09-134-001C-5542	Sequence 5542, Ap
273	6	0.8	261	4	US-09-036-987A-29	Sequence 29, Appl	346	6	0.8	329	4	US-09-651-200-19	Sequence 19, Appl
274	6	0.8	261	4	US-09-370-700-29	Sequence 29, Appl	347	6	0.8	330	4	US-09-267-031-12	Sequence 12, Appl
275	6	0.8	262	2	US-08-038-761A-1	Sequence 1, Appl	348	6	0.8	332	5	PCT-US94-14277-2	Sequence 2, Appl
276	6	0.8	263	5	PCT-US91-06532-2	Sequence 2, Appl	349	6	0.8	333	3	US-08-938-546-2	Sequence 2, Appl
277	6	0.8	264	4	US-08-483-533-40	Sequence 40, Appl	350	6	0.8	333	3	US-09-340-812-2	Sequence 2, Appl
278	6	0.8	264	4	US-09-283-471A-40	Sequence 40, Appl	351	6	0.8	333	3	US-09-095-117-2	Sequence 2, Appl
279	6	0.8	264	4	US-08-936-165A-477	Sequence 477, App	352	6	0.8	335	4	US-09-095-117-4	Sequence 4, Appl
280	6	0.8	265	4	US-09-199-637A-31	Sequence 31, Appl	353	6	0.8	336	1	US-07-667-276A-8	Sequence 8, Appl
281	6	0.8	265	4	US-09-144-776B-14	Sequence 14, Appl	354	6	0.8	336	1	US-08-118-270-60	Sequence 54, Appl
282	6	0.8	267	4	US-08-430-225A-20	Sequence 20, Appl	355	6	0.8	336	4	US-09-105-390-60	Sequence 60, Appl
283	6	0.8	268	1	US-08-270-584A-2	Sequence 2, Appl	356	6	0.8	336	5	PCT-US93-08528-54	Sequence 54, Appl
284	6	0.8	268	1	US-08-568-031-2	Sequence 2, Appl	357	6	0.8	337	4	US-09-188-930-186	Sequence 186, App
285	6	0.8	268	2	US-08-765-192-2	Sequence 2, Appl	358	6	0.8	339	3	US-08-938-546-4	Sequence 4, Appl
286	6	0.8	268	2	US-08-966-319-2	Sequence 2, Appl	359	6	0.8	339	3	US-08-938-546-4	Sequence 4, Appl
287	6	0.8	268	3	US-09-153-304-2	Sequence 2, Appl	360	6	0.8	343	2	US-09-340-812-4	Sequence 4, Appl
288	6	0.8	268	3	US-09-159-793-2	Sequence 2, Appl	361	6	0.8	343	2	US-08-788-539A-2	Sequence 2, Appl
289	6	0.8	268	4	US-09-430-564-12	Sequence 12, Appl	362	6	0.8	349	2	US-08-558-823-4	Sequence 4, Appl
290	6	0.8	269	1	US-08-447-554-5	Sequence 5, Appl	363	6	0.8	351	1	US-08-468-847B-16	Sequence 16, Appl
291	6	0.8	269	1	US-08-468-763-17	Sequence 17, Appl	364	6	0.8	352	4	US-09-504-358-18	Sequence 18, Appl
292	6	0.8	269	1	US-08-448-160-5	Sequence 5, Appl	365	6	0.8	353	1	US-09-954-314-18	Sequence 18, Appl
293	6	0.8	269	2	US-08-393-996A-17	Sequence 17, Appl	366	6	0.8	353	1	US-08-097-831-6	Sequence 6, Appl
294	6	0.8	274	4	US-09-250-677-2	Sequence 2, Appl	367	6	0.8	353	1	US-08-034-650-11	Sequence 11, Appl
295	6	0.8	277	4	US-09-198-119C-61	Sequence 61, Appl	368	6	0.8	353	4	US-08-449-015-11	Sequence 11, Appl
296	6	0.8	279	4	US-09-010-999-2	Sequence 2, Appl	369	6	0.8	357	4	US-09-647-540A-4	Sequence 4, Appl
297	6	0.8	279	4	US-09-134-001C-4878	Sequence 4878, Ap	370	6	0.8	358	4	US-09-134-001C-4405	Sequence 4405, Ap
298	6	0.8	279	4	US-09-198-119C-51	Sequence 51, Appl	371	6	0.8	358	4	US-09-626-410-7	Sequence 7, Appl
299	6	0.8	280	4	US-09-198-119C-77	Sequence 77, Appl	372	6	0.8	358	4	US-09-626-410-8	Sequence 8, Appl
300	6	0.8	283	4	US-09-198-119C-63	Sequence 63, Appl	373	6	0.8	358	4	US-09-626-410-9	Sequence 9, Appl
301	6	0.8	284	4	US-09-056-556-162	Sequence 162, App	374	6	0.8	358	4	US-09-626-410-10	Sequence 10, Appl
302	6	0.8	284	4	US-09-072-596-157	Sequence 157, App	375	6	0.8	358	4	US-09-626-410-11	Sequence 11, Appl
303	6	0.8	288	4	US-09-095-758-8	Sequence 8, Appl	376	6	0.8	358	4	US-09-626-410-12	Sequence 12, Appl
304	6	0.8	288	4	US-09-422-968-8	Sequence 8, Appl	377	6	0.8	358	4	US-09-626-410-14	Sequence 14, Appl
305	6	0.8	288	4	US-09-708-015A-8	Sequence 8, Appl	378	6	0.8	358	4	US-09-116-188-7	Sequence 7, Appl
306	6	0.8	290	2	US-08-614-686A-1	Sequence 1, Appl	379	6	0.8	358	4	US-09-116-188-8	Sequence 8, Appl
307	6	0.8	293	4	US-08-438-745-4	Sequence 4, Appl	380	6	0.8	358	4	US-09-116-188-9	Sequence 9, Appl
308	6	0.8	293	4	US-08-438-745-6	Sequence 6, Appl	381	6	0.8	358	4	US-09-116-188-10	Sequence 10, Appl
309	6	0.8	293	4	US-09-219-019-4	Sequence 4, Appl	382	6	0.8	358	4	US-09-116-188-11	Sequence 11, Appl
310	6	0.8	293	4	US-09-219-019-6	Sequence 6, Appl	383	6	0.8	358	4	US-09-116-188-12	Sequence 12, Appl
311	6	0.8	293	5	PCT-US94-05669A-4	Sequence 4, Appl	384	6	0.8	358	4	US-09-116-188-14	Sequence 14, Appl
312	6	0.8	293	5	PCT-US94-05669A-6	Sequence 6, Appl	385	6	0.8	358	4	US-09-626-047-7	Sequence 7, Appl
313	6	0.8	297	1	US-08-534-910B-6	Sequence 6, Appl	386	6	0.8	358	4	US-09-626-047-8	Sequence 8, Appl
314	6	0.8	297	1	US-08-534-910B-7	Sequence 7, Appl	387	6	0.8	358	4	US-09-626-047-9	Sequence 9, Appl
315	6	0.8	297	1	US-08-534-910B-8	Sequence 8, Appl	388	6	0.8	358	4	US-09-626-047-10	Sequence 10, Appl
316	6	0.8	297	1	US-08-534-910B-9	Sequence 9, Appl	389	6	0.8	358	4	US-09-626-047-11	Sequence 11, Appl
317	6	0.8	297	1	US-08-534-910B-10	Sequence 10, Appl	390	6	0.8	358	4	US-09-626-047-12	Sequence 12, Appl
318	6	0.8	297	3	US-08-886-466-2	Sequence 2, Appl	391	6	0.8	358	4	US-09-626-047-14	Sequence 14, Appl
319	6	0.8	297	4	US-09-475-304-2	Sequence 2, Appl	392	6	0.8	358	4	US-09-626-343-8	Sequence 8, Appl

393	6	0.8	358	4	US-09-626-343-9	Sequence 9, Appli	466	6	0.8	430	5	PCT-US93-05446-9	Sequence 9, Appli
394	6	0.8	358	4	US-09-626-343-10	Sequence 10, Appl	467	6	0.8	430	5	PCT-US93-07189-4	Sequence 4, Appli
395	6	0.8	358	4	US-09-626-343-11	Sequence 11, Appl	468	6	0.8	430	5	PCT-US93-07190-19	Sequence 19, Appl
396	6	0.8	358	4	US-09-626-343-12	Sequence 12, Appl	469	6	0.8	430	5	PCT-US93-07231-19	Sequence 19, Appl
397	6	0.8	358	4	US-09-626-343-14	Sequence 14, Appl	470	6	0.8	430	5	PCT-US93-08742-19	Sequence 19, Appl
398	6	0.8	358	4	US-09-354-922-8	Sequence 8, Appli	471	6	0.8	430	5	PCT-US93-08808-19	Sequence 19, Appl
399	6	0.8	358	4	US-09-354-922-9	Sequence 9, Appli	472	6	0.8	430	5	PCT-US93-08885-19	Sequence 19, Appl
400	6	0.8	358	4	US-09-354-922-10	Sequence 10, Appl	473	6	0.8	430	5	PCT-US93-10520-6	Sequence 6, Appli
401	6	0.8	358	4	US-09-354-922-11	Sequence 11, Appl	474	6	0.8	432	4	US-09-403-089A-1	Sequence 1, Appli
402	6	0.8	358	4	US-09-354-922-12	Sequence 12, Appl	475	6	0.8	434	1	US-07-923-692C-6	Sequence 6, Appli
403	6	0.8	358	4	US-09-354-922-13	Sequence 13, Appl	476	6	0.8	434	1	US-08-184-237-6	Sequence 6, Appli
404	6	0.8	358	4	US-09-354-922-15	Sequence 15, Appl	477	6	0.8	434	2	US-08-482-920-6	Sequence 6, Appli
405	6	0.8	358	2	US-08-846-762-81	Sequence 81, Appl	478	6	0.8	434	3	US-08-484-341-6	Sequence 6, Appli
406	6	0.8	372	1	US-07-813-584A-3	Sequence 3, Appli	479	6	0.8	434	4	US-08-483-502-6	Sequence 6, Appli
407	6	0.8	373	1	US-08-330-515-3	Sequence 3, Appli	480	6	0.8	434	4	US-09-726-651A-6	Sequence 6, Appli
408	6	0.8	373	3	US-08-746-883-4	Sequence 4, Appli	481	6	0.8	436	4	US-09-134-001C-4295	Sequence 4295, Ap
409	6	0.8	376	6	5180810-1	Patent No. 5180810	482	6	0.8	438	1	US-08-111-939-17	Sequence 17, Appl
410	6	0.8	391	4	US-09-134-001C-4880	Sequence 4880, Ap	483	6	0.8	438	3	US-09-187-331-2	Sequence 2, Appli
411	6	0.8	392	1	US-07-994-133-2	Sequence 2, Appli	484	6	0.8	438	4	US-09-470-946-2	Sequence 2, Appli
412	6	0.8	392	6	5196304-2	Patent No. 5196304	485	6	0.8	440	4	US-09-077-955-26	Sequence 26, Appl
413	6	0.8	394	3	US-08-673-814-6	Sequence 6, Appli	486	6	0.8	443	4	US-08-660-765A-4	Sequence 4, Appli
414	6	0.8	394	4	US-09-115-824-6	Sequence 6, Appli	487	6	0.8	449	2	US-08-927-394-2	Sequence 2, Appli
415	6	0.8	396	4	US-08-860-656B-4	Sequence 4, Appli	488	6	0.8	450	2	US-08-861-464-2	Sequence 2, Appli
416	6	0.8	402	4	US-08-861-774B-14	Sequence 14, Appl	489	6	0.8	450	2	US-08-396-001-2	Sequence 2, Appli
417	6	0.8	406	3	US-08-334-494-2	Sequence 2, Appli	490	6	0.8	450	4	US-09-323-433A-2	Sequence 2, Appli
418	6	0.8	406	3	US-09-143-068-2	Sequence 2, Appli	491	6	0.8	453	4	US-09-171-461-48	Sequence 48, Appl
419	6	0.8	406	4	US-08-861-774B-25	Sequence 25, Appl	492	6	0.8	456	4	US-09-077-955-25	Sequence 25, Appl
420	6	0.8	406	4	US-09-143-707-2	Sequence 2, Appli	493	6	0.8	458	4	US-09-233-989-9	Sequence 9, Appli
421	6	0.8	406	4	US-09-202-089-2	Sequence 2, Appli	494	6	0.8	461	2	US-08-463-587A-26	Sequence 26, Appl
422	6	0.8	406	4	US-09-511-133-2	Sequence 2, Appli	495	6	0.8	461	3	US-08-463-667A-4	Sequence 4, Appli
423	6	0.8	406	4	US-09-690-169-2	Sequence 2, Appli	496	6	0.8	461	3	US-08-923-854-26	Sequence 26, Appl
424	6	0.8	406	4	US-09-511-631-2	Sequence 2, Appli	497	6	0.8	461	5	PCT-US91-09133-27	Sequence 27, Appl
425	6	0.8	409	2	US-08-576-626A-48	Sequence 48, Appl	498	6	0.8	462	4	US-09-189-527-13	Sequence 13, Appl
426	6	0.8	417	2	US-08-682-847-6	Sequence 6, Appli	499	6	0.8	465	4	US-08-845-258-30	Sequence 30, Appl
427	6	0.8	417	4	US-09-063-676-2	Sequence 2, Appli	500	6	0.8	465	4	US-08-990-571-30	Sequence 30, Appl
428	6	0.8	427	3	US-08-705-771-16	Sequence 16, Appl	501	6	0.8	465	4	US-08-723-142A-30	Sequence 30, Appl
429	6	0.8	427	4	US-08-065-844A-2	Sequence 2, Appli	502	6	0.8	465	4	US-09-528-784A-30	Sequence 30, Appl
430	6	0.8	428	1	US-07-973-324A-4	Sequence 4, Appli	503	6	0.8	468	4	US-09-149-476-387	Sequence 387, App
431	6	0.8	428	1	US-08-343-380-4	Sequence 4, Appli	504	6	0.8	471	4	US-09-004-838-45	Sequence 45, Appl
432	6	0.8	428	4	US-09-072-435-4	Sequence 4, Appli	505	6	0.8	472	4	US-09-625-188-6	Sequence 6, Appli
433	6	0.8	428	4	US-09-072-917A-4	Sequence 4, Appli	506	6	0.8	475	4	US-09-004-838-97	Sequence 97, Appl
434	6	0.8	429	1	US-08-745-977-4	Sequence 4, Appli	507	6	0.8	477	4	US-09-004-838-45	Sequence 45, Appl
435	6	0.8	429	3	US-09-040-699A-4	Sequence 4, Appli	508	6	0.8	477	4	US-09-004-838-99	Sequence 99, Appl
436	6	0.8	430	1	US-07-841-646-25	Sequence 25, Appl	509	6	0.8	482	1	US-08-358-160-5	Sequence 5, Appli
437	6	0.8	430	1	US-07-901-703-9	Sequence 9, Appli	510	6	0.8	482	4	US-09-430-564-2	Sequence 2, Appli
438	6	0.8	430	1	US-08-147-023-25	Sequence 25, Appl	511	6	0.8	484	1	US-08-358-160-7	Sequence 7, Appli
439	6	0.8	430	1	US-08-278-729A-19	Sequence 19, Appl	512	6	0.8	486	4	US-09-134-001C-3593	Sequence 3593, Ap
440	6	0.8	430	1	US-08-480-528A-6	Sequence 6, Appli	513	6	0.8	491	2	US-08-687-916-23	Sequence 23, Appl
441	6	0.8	430	1	US-08-479-666-6	Sequence 6, Appli	514	6	0.8	491	4	US-09-138-614-23	Sequence 23, Appl
442	6	0.8	430	1	US-08-155-343A-19	Sequence 19, Appl	515	6	0.8	492	1	US-07-783-705A-4	Sequence 4, Appli
443	6	0.8	430	1	US-08-406-672-19	Sequence 19, Appl	516	6	0.8	492	2	US-08-644-271-32	Sequence 32, Appl
444	6	0.8	430	1	US-08-643-563A-19	Sequence 19, Appl	517	6	0.8	492	4	US-09-077-955-36	Sequence 36, Appl
445	6	0.8	430	1	US-08-447-570-25	Sequence 25, Appl	518	6	0.8	495	2	US-08-687-916-22	Sequence 22, Appl
446	6	0.8	430	1	US-08-643-763A-19	Sequence 19, Appl	519	6	0.8	495	4	US-09-138-614-22	Sequence 22, Appl
447	6	0.8	430	1	US-08-462-623-19	Sequence 19, Appl	520	6	0.8	495	4	US-09-430-564-3	Sequence 3, Appli
448	6	0.8	430	1	US-08-451-953A-19	Sequence 19, Appl	521	6	0.8	495	6	5258283-9	Patent No. 5258283
449	6	0.8	430	2	US-08-459-346-4	Sequence 4, Appli	522	6	0.8	496	3	US-08-881-784-1	Sequence 1, Appli
450	6	0.8	430	2	US-08-445-468A-19	Sequence 19, Appl	523	6	0.8	496	4	US-09-292-768-2	Sequence 2, Appli
451	6	0.8	430	2	US-08-901-200A-6	Sequence 6, Appli	524	6	0.8	496	4	US-09-292-768-64	Sequence 64, Appl
452	6	0.8	430	2	US-08-449-700-25	Sequence 25, Appl	525	6	0.8	496	4	US-09-292-768-66	Sequence 66, Appl
453	6	0.8	430	2	US-08-449-699A-25	Sequence 25, Appl	526	6	0.8	496	4	US-09-172-339-6	Sequence 6, Appli
454	6	0.8	430	2	US-08-461-397A-19	Sequence 19, Appl	527	6	0.8	510	1	US-08-220-677A-2	Sequence 2, Appli
455	6	0.8	430	2	US-08-912-088-19	Sequence 19, Appl	528	6	0.8	511	3	US-08-464-052-2	Sequence 2, Appli
456	6	0.8	430	3	US-08-378-730A-19	Sequence 19, Appl	529	6	0.8	511	4	US-08-461-002-2	Sequence 2, Appli
457	6	0.8	430	3	US-08-889-419-4	Sequence 4, Appli	530	6	0.8	511	5	US-08-689-411-2	Sequence 2, Appli
458	6	0.8	430	3	US-08-445-467-19	Sequence 19, Appl	531	6	0.8	511	5	PCT-US94-09863-2	Sequence 2, Appli
459	6	0.8	430	3	US-08-480-515A-19	Sequence 19, Appl	532	6	0.8	522	6	RE34606-6	Patent No. RE34, 60
460	6	0.8	430	4	US-09-219-391-6	Sequence 6, Appli	533	6	0.8	524	4	US-09-230-388-1	Sequence 1, Appli
461	6	0.8	430	4	US-09-170-936-19	Sequence 19, Appl	534	6	0.8	525	2	US-08-749-902-7	Sequence 7, Appli
462	6	0.8	430	4	US-08-402-542-4	Sequence 4, Appli	535	6	0.8	525	2	US-08-749-902-8	Sequence 8, Appli
463	6	0.8	430	4	US-08-461-113-19	Sequence 19, Appl	536	6	0.8	525	4	US-08-543-634-28	Sequence 28, Appl
464	6	0.8	430	5	PCT-US90-0765A-2	Sequence 2, Appli	537	6	0.8	525	4	US-09-430-564-16	Sequence 16, Appl
465	6	0.8	430	5	PCT-US92-01968-19	Sequence 19, Appl	538	6	0.8	525	5	PCT-US95-13703-28	Sequence 28, Appl

539	6	0.8	527	3	US-08-907-229-2	Sequence 2, Appli	612	6	0.8	678	5	PCT-US93-03027-3	Sequence 3, Appli
540	6	0.8	527	4	US-09-574-462-2	Sequence 2, Appli	613	6	0.8	680	1	US-08-211-430-2	Sequence 2, Appli
541	6	0.8	527	4	US-08-677-970-3	Sequence 2, Appli	614	6	0.8	680	3	US-08-761-136-1	Sequence 1, Appli
542	6	0.8	529	4	US-09-291-922-28	Sequence 28, Appli	615	6	0.8	694	1	US-08-171-382-2	Sequence 2, Appli
543	6	0.8	540	4	US-08-542-634-26	Sequence 26, Appli	616	6	0.8	694	1	US-08-309-420-2	Sequence 2, Appli
544	6	0.8	540	5	PCT-US95-13703-26	Sequence 26, Appli	617	6	0.8	694	1	US-08-309-419-2	Sequence 2, Appli
545	6	0.8	541	3	US-09-058-489-40	Sequence 40, Appli	618	6	0.8	694	4	US-09-294-531B-4	Sequence 4, Appli
546	6	0.8	545	4	US-09-351-224E-3	Sequence 3, Appli	619	6	0.8	694	5	PCT-US95-11856-2	Sequence 2, Appli
547	6	0.8	548	4	US-08-477-292-16	Sequence 16, Appli	620	6	0.8	700	4	PCT-US95-11878-2	Sequence 2, Appli
548	6	0.8	549	4	US-08-542-634-16	Sequence 16, Appli	621	6	0.8	700	2	US-08-568-459A-10	Sequence 10, Appli
549	6	0.8	549	5	PCT-US95-13703-16	Sequence 16, Appli	622	6	0.8	700	2	US-08-487-826B-10	Sequence 10, Appli
550	6	0.8	555	2	US-08-982-232-7	Sequence 7, Appli	623	6	0.8	700	4	US-09-210-288-10	Sequence 10, Appli
551	6	0.8	555	2	US-08-982-232-14	Sequence 14, Appli	624	6	0.8	701	3	US-09-087-727-2	Sequence 2, Appli
552	6	0.8	563	1	US-08-442-884-1	Sequence 1, Appli	625	6	0.8	721	4	US-09-146-221-6	Sequence 4, Appli
553	6	0.8	568	4	US-08-939-309-2	Sequence 2, Appli	626	6	0.8	724	4	US-09-146-221-6	Sequence 4, Appli
554	6	0.8	575	4	US-09-171-461-30	Sequence 30, Appli	627	6	0.8	727	4	US-09-146-221-6	Sequence 8, Appli
555	6	0.8	582	1	US-08-261-086-2	Sequence 2, Appli	628	6	0.8	729	4	US-09-114-001C-4728	Sequence 8, Appli
556	6	0.8	585	1	US-08-142-439A-6	Sequence 6, Appli	629	6	0.8	738	3	US-08-864-038A-3	Sequence 4728, Ap
557	6	0.8	585	2	US-08-142-551B-125	Sequence 125, App	630	6	0.8	738	6	US-08-864-038A-3	Sequence 3, Appli
558	6	0.8	585	2	US-08-869-477-6	Sequence 6, Appli	631	6	0.8	740	6	5262177-2	Sequence 3, Appli
559	6	0.8	585	2	US-08-468-249A-19	Sequence 19, Appli	632	6	0.8	740	5	PCT-US92-08756A-10	Sequence 10, Appli
560	6	0.8	587	4	US-09-147-923-2	Sequence 2, Appli	633	6	0.8	745	2	US-09-010-928B-28	Sequence 28, Appli
561	6	0.8	593	1	US-07-668-648-4	Sequence 4, Appli	634	6	0.8	748	2	US-08-997-080-154	Sequence 154, App
562	6	0.8	593	2	US-08-429-998-4	Sequence 4, Appli	635	6	0.8	748	2	US-08-997-362-154	Sequence 154, App
563	6	0.8	593	2	US-08-431-333-4	Sequence 4, Appli	636	6	0.8	748	2	US-08-997-362-154	Sequence 154, App
564	6	0.8	593	4	US-08-931-862-17	Sequence 17, Appli	637	6	0.8	748	4	US-09-095-855-154	Sequence 154, App
565	6	0.8	593	5	PCT-US91-02321-17	Sequence 17, Appli	638	6	0.8	748	4	US-09-324-542-154	Sequence 154, App
566	6	0.8	598	4	US-09-310-463-10	Sequence 10, Appli	639	6	0.8	752	1	US-08-309-512-8	Sequence 8, Appli
567	6	0.8	598	4	US-08-842-248A-10	Sequence 10, Appli	640	6	0.8	752	5	PCT-US92-08756A-8	Sequence 8, Appli
568	6	0.8	602	4	US-09-374-454-19	Sequence 19, Appli	641	6	0.8	754	3	US-09-005-180A-3	Sequence 3, Appli
569	6	0.8	613	2	US-08-915-207-2	Sequence 2, Appli	642	6	0.8	764	4	US-07-757-342D-5	Sequence 5, Appli
570	6	0.8	613	2	US-08-915-207-4	Sequence 4, Appli	643	6	0.8	764	4	US-07-741-453A-54	Sequence 54, Appli
571	6	0.8	613	4	US-09-238-555-2	Sequence 2, Appli	644	6	0.8	764	4	US-07-741-453A-59	Sequence 59, Appli
572	6	0.8	613	4	US-09-238-555-4	Sequence 4, Appli	645	6	0.8	764	4	US-07-741-453A-61	Sequence 61, Appli
573	6	0.8	615	4	US-08-985-950-16	Sequence 16, Appli	646	6	0.8	765	1	US-08-309-512-5	Sequence 5, Appli
574	6	0.8	615	4	US-08-985-950-18	Sequence 18, Appli	647	6	0.8	765	5	PCT-US92-08756A-5	Sequence 5, Appli
575	6	0.8	616	4	US-09-134-001C-4618	Sequence 4618, Ap	648	6	0.8	771	4	US-09-090-793-8	Sequence 8, Appli
576	6	0.8	618	4	US-08-961-083-72	Sequence 72, Appli	649	6	0.8	780	1	US-08-375-709-13	Sequence 13, Appli
577	6	0.8	619	4	US-09-233-200-37	Sequence 37, Appli	650	6	0.8	780	1	US-08-752-929-13	Sequence 13, Appli
578	6	0.8	619	4	US-09-233-200-40	Sequence 40, Appli	651	6	0.8	796	3	US-09-005-180A-1	Sequence 1, Appli
579	6	0.8	619	4	US-09-233-200-57	Sequence 57, Appli	652	6	0.8	796	4	US-08-868-639A-2	Sequence 2, Appli
580	6	0.8	619	4	US-09-233-197-37	Sequence 37, Appli	653	6	0.8	796	4	US-09-757-014-2	Sequence 2, Appli
581	6	0.8	619	4	US-09-233-197-40	Sequence 40, Appli	654	6	0.8	806	4	US-09-134-001C-4314	Sequence 4314, Ap
582	6	0.8	619	4	US-09-233-197-57	Sequence 57, Appli	655	6	0.8	822	4	US-09-219-849-59	Sequence 49, Appli
583	6	0.8	619	4	US-09-233-201-37	Sequence 37, Appli	656	6	0.8	823	1	US-08-461-551-2	Sequence 2, Appli
584	6	0.8	619	4	US-09-233-201-40	Sequence 40, Appli	657	6	0.8	823	4	US-08-461-551-2	Sequence 2, Appli
585	6	0.8	619	4	US-09-233-201-57	Sequence 57, Appli	658	6	0.8	840	2	US-08-467-822-55	Sequence 25, Appli
586	6	0.8	623	4	US-08-975-762-68	Sequence 68, Appli	659	6	0.8	840	4	US-08-432-697-25	Sequence 25, Appli
587	6	0.8	623	4	US-09-295-028-68	Sequence 68, Appli	660	6	0.8	840	4	US-08-466-248-55	Sequence 25, Appli
588	6	0.8	623	4	US-09-106-582-68	Sequence 68, Appli	661	6	0.8	843	2	US-08-867-129-2	Sequence 22, Appli
589	6	0.8	626	2	US-08-596-300A-7	Sequence 7, Appli	662	6	0.8	843	3	US-09-215-966-12	Sequence 22, Appli
590	6	0.8	626	2	US-08-596-300A-14	Sequence 14, Appli	663	6	0.8	856	4	US-09-134-001C-5438	Sequence 5338, Ap
591	6	0.8	630	3	US-08-771-986A-2	Sequence 2, Appli	664	6	0.8	862	2	US-08-685-118-2	Sequence 2, Appli
592	6	0.8	630	3	US-08-769-802A-2	Sequence 2, Appli	665	6	0.8	862	2	US-08-915-495-2	Sequence 2, Appli
593	6	0.8	642	4	US-08-461-562B-9	Sequence 9, Appli	666	6	0.8	862	2	US-08-914-520-2	Sequence 2, Appli
594	6	0.8	650	4	US-09-134-001C-4444	Sequence 4444, Ap	667	6	0.8	865	4	US-08-842-382-2	Sequence 2, Appli
595	6	0.8	658	1	US-08-409-995-5	Sequence 5, Appli	668	6	0.8	869	2	US-08-483-101-15	Sequence 15, Appli
596	6	0.8	658	3	US-08-685-467-5	Sequence 5, Appli	669	6	0.8	870	2	US-09-010-928B-2	Sequence 2, Appli
597	6	0.8	658	4	US-08-913-942-5	Sequence 5, Appli	670	6	0.8	893	4	US-09-514-302-4	Sequence 4, Appli
598	6	0.8	659	1	US-08-240-049B-16	Sequence 16, Appli	671	6	0.8	902	1	US-08-396-479B-6	Sequence 6, Appli
599	6	0.8	659	1	US-08-259-148A-20	Sequence 20, Appli	672	6	0.8	902	1	US-08-818-823-6	Sequence 6, Appli
600	6	0.8	659	1	US-08-484-054-20	Sequence 20, Appli	673	6	0.8	910	5	US-09-228-986-72	Sequence 72, Appli
601	6	0.8	659	2	US-07-876-941A-20	Sequence 20, Appli	674	6	0.8	912	4	PCT-US95-03747-2	Sequence 2, Appli
602	6	0.8	659	4	US-08-477-292-14	Sequence 14, Appli	675	6	0.8	968	4	US-09-228-986-76	Sequence 76, Appli
603	6	0.8	659	4	US-07-870-985A-20	Sequence 20, Appli	676	6	0.8	983	4	US-09-134-001C-3814	Sequence 3814, Ap
604	6	0.8	660	4	US-08-542-634-14	Sequence 14, Appli	677	6	0.8	987	1	US-08-436-044-6	Sequence 6, Appli
605	6	0.8	660	4	US-09-462-606-48	Sequence 48, Appli	678	6	0.8	987	5	US-08-436-054-6	Sequence 6, Appli
606	6	0.8	660	5	US-09-462-606-57	Sequence 57, Appli	679	6	0.8	987	5	PCT-US95-08812-6	Sequence 6, Appli
607	6	0.8	660	5	PCT-US95-13703-14	Sequence 14, Appli	680	6	0.8	994	3	US-08-699-103B-3	Sequence 3, Appli
608	6	0.8	661	3	US-08-394-1326-2	Sequence 2, Appli	681	6	0.8	994	4	US-09-229-059-3	Sequence 3, Appli
609	6	0.8	661	3	US-09-082-306-2	Sequence 2, Appli	682	6	0.8	995	5	PCT-US95-04910-14	Sequence 14, Appli
610	6	0.8	662	4	US-09-061-768A-25	Sequence 25, Appli	683	6	0.8	1040	4	US-09-134-001C-5365	Sequence 5365, Ap
611	6	0.8	667	4	US-08-959-004-11	Sequence 11, Appli	684	6	0.8	1057	3	US-08-931-820-1	Sequence 1, Appli

585	6	0.8	1059	4	US-09-697-367-16	Sequence 16, Appl	758	1189	4	US-09-251-885-4	Sequence 4, Appl
586	6	0.8	1060	3	US-08-931-820-3	Sequence 3, Appl	759	1189	4	US-09-251-885-6	Sequence 6, Appl
587	6	0.8	1066	2	US-08-633-770A-1	Sequence 1, Appl	760	1189	4	US-09-251-885-8	Sequence 8, Appl
588	6	0.8	1066	4	US-09-004-838-24	Sequence 24, Appl	761	1189	4	US-09-251-885-10	Sequence 10, Appl
589	6	0.8	1070	2	US-08-633-770A-2	Sequence 2, Appl	762	1189	4	US-09-251-885-12	Sequence 12, Appl
590	6	0.8	1079	1	US-08-455-543A-55	Sequence 55, Appl	763	1189	4	US-09-001-982-2	Sequence 2, Appl
591	6	0.8	1079	2	US-08-223-305C-55	Sequence 55, Appl	764	1189	4	US-09-337-635-2	Sequence 2, Appl
592	6	0.8	1084	1	US-08-455-543A-56	Sequence 56, Appl	765	1189	4	US-09-337-635-4	Sequence 4, Appl
593	6	0.8	1084	2	US-08-223-305C-56	Sequence 56, Appl	766	1189	4	US-09-337-635-6	Sequence 6, Appl
594	6	0.8	1086	1	US-08-455-543A-54	Sequence 54, Appl	767	1189	4	US-09-337-635-8	Sequence 8, Appl
595	6	0.8	1086	2	US-08-223-305C-54	Sequence 54, Appl	768	1189	4	US-09-337-635-10	Sequence 10, Appl
596	6	0.8	1086	6	5386025-8	Patent No. 5386025	769	1189	4	US-09-337-635-12	Sequence 12, Appl
597	6	0.8	1091	1	US-07-745-206A-25	Sequence 25, Appl	770	1189	4	US-09-337-635-59	Sequence 59, Appl
598	6	0.8	1091	1	US-08-455-543A-52	Sequence 52, Appl	771	1189	4	US-09-337-635-61	Sequence 61, Appl
599	6	0.8	1091	2	US-08-223-305C-52	Sequence 52, Appl	772	1189	4	US-09-337-280-2	Sequence 2, Appl
600	6	0.8	1091	2	US-08-311-363-25	Sequence 25, Appl	773	1189	4	US-09-337-280-4	Sequence 4, Appl
601	6	0.8	1091	3	US-08-713-118-4	Sequence 4, Appl	774	1189	4	US-09-337-280-6	Sequence 6, Appl
602	6	0.8	1091	4	US-09-452-007-4	Sequence 4, Appl	775	1189	4	US-09-337-280-8	Sequence 8, Appl
603	6	0.8	1098	1	US-08-409-995-2	Sequence 2, Appl	776	1189	4	US-09-337-280-10	Sequence 10, Appl
604	6	0.8	1098	3	US-08-685-467-2	Sequence 2, Appl	777	1189	4	US-09-337-280-12	Sequence 12, Appl
605	6	0.8	1098	4	US-09-377-155-32	Sequence 32, Appl	778	1189	4	US-09-337-280-59	Sequence 59, Appl
606	6	0.8	1098	4	US-08-913-942-2	Sequence 2, Appl	779	1189	4	US-09-337-280-61	Sequence 61, Appl
607	6	0.8	1098	4	US-09-669-974-32	Sequence 32, Appl	780	1189	5	PCT-US92-11337-16	Sequence 16, Appl
608	6	0.8	1098	4	US-09-268-347-44	Sequence 44, Appl	781	1189	6	5188960-6	Patent No. 5188960
609	6	0.8	1103	1	US-08-455-543A-53	Sequence 53, Appl	782	1189	6	US-08-239-474A-12	Sequence 12, Appl
610	6	0.8	1103	2	US-08-223-305C-53	Sequence 53, Appl	783	1190	2	US-08-732-495-12	Sequence 12, Appl
611	6	0.8	1103	3	US-09-162-373-1	Sequence 1, Appl	784	1205	1	US-07-908-245-2	Sequence 2, Appl
612	6	0.8	1103	4	US-09-467-946-1	Sequence 1, Appl	785	1205	2	US-08-319-866-10	Sequence 10, Appl
613	6	0.8	1106	1	US-08-435-675B-5	Sequence 5, Appl	786	1205	4	US-09-123-708-6	Sequence 6, Appl
614	6	0.8	1106	1	US-08-336-257A-8	Sequence 8, Appl	787	1205	4	US-09-123-624-6	Sequence 6, Appl
615	6	0.8	1132	4	US-09-328-784A-87	Sequence 87, Appl	788	1213	4	US-09-413-814-79	Sequence 79, Appl
616	6	0.8	1140	4	US-09-651-656-21	Sequence 21, Appl	789	1237	1	US-08-241-853-2	Sequence 2, Appl
617	6	0.8	1140	4	US-09-650-855-21	Sequence 21, Appl	790	1237	1	US-08-850-917-2	Sequence 2, Appl
618	6	0.8	1149	1	US-07-915-203-2	Sequence 2, Appl	791	1276	1	US-08-222-616-24	Sequence 24, Appl
619	6	0.8	1149	1	US-08-272-887-2	Sequence 2, Appl	792	1276	4	US-08-446-648-24	Sequence 24, Appl
620	6	0.8	1149	2	US-08-789-449-2	Sequence 2, Appl	793	1276	5	PCT-US95-04228-24	Sequence 24, Appl
621	6	0.8	1163	1	US-08-239-474A-11	Sequence 11, Appl	794	1286	6	5206163-1	Patent No. 5206163
622	6	0.8	1163	2	US-08-732-495-11	Sequence 11, Appl	795	1306	4	US-08-999-774A-13	Sequence 13, Appl
623	6	0.8	1163	4	US-09-178-252-21	Sequence 21, Appl	796	1323	4	US-09-004-838-90	Sequence 90, Appl
624	6	0.8	1189	1	US-07-828-788A-16	Sequence 16, Appl	797	1341	3	US-08-963-825-18	Sequence 18, Appl
625	6	0.8	1189	1	US-08-356-034-6	Sequence 6, Appl	798	1341	4	US-09-500-811-18	Sequence 18, Appl
626	6	0.8	1189	2	US-08-602-737-2	Sequence 2, Appl	799	1341	4	US-09-570-573-18	Sequence 18, Appl
627	6	0.8	1189	2	US-08-980-071-2	Sequence 2, Appl	800	1341	4	US-09-548-608-18	Sequence 18, Appl
628	6	0.8	1189	2	US-08-980-071-4	Sequence 4, Appl	801	1382	3	US-09-057-570-4	Sequence 4, Appl
629	6	0.8	1189	2	US-08-980-071-6	Sequence 6, Appl	802	1418	3	US-08-963-825-20	Sequence 20, Appl
630	6	0.8	1189	2	US-08-980-071-8	Sequence 8, Appl	803	1418	4	US-09-010-999-1	Sequence 1, Appl
631	6	0.8	1189	2	US-08-980-071-10	Sequence 10, Appl	804	1418	4	US-09-500-811-20	Sequence 20, Appl
632	6	0.8	1189	2	US-08-980-071-12	Sequence 12, Appl	805	1418	4	US-09-570-573-20	Sequence 20, Appl
633	6	0.8	1189	2	US-08-980-071-59	Sequence 59, Appl	806	1418	4	US-09-548-608-20	Sequence 20, Appl
634	6	0.8	1189	2	US-08-980-071-61	Sequence 61, Appl	807	1442	4	US-08-316-650-12	Sequence 12, Appl
635	6	0.8	1189	2	US-08-757-536-2	Sequence 2, Appl	808	1442	5	PCT-US95-02251-12	Sequence 12, Appl
636	6	0.8	1189	2	US-08-757-536-4	Sequence 4, Appl	809	1452	2	US-08-653-971-4	Sequence 4, Appl
637	6	0.8	1189	2	US-08-757-536-6	Sequence 6, Appl	810	1452	2	US-08-991-258A-4	Sequence 4, Appl
638	6	0.8	1189	2	US-08-757-536-8	Sequence 8, Appl	811	1452	2	US-08-763-399-4	Sequence 4, Appl
639	6	0.8	1189	2	US-08-757-536-10	Sequence 10, Appl	812	1452	3	US-08-991-953A-4	Sequence 4, Appl
640	6	0.8	1189	2	US-08-757-536-12	Sequence 12, Appl	813	1461	4	US-09-585-887-9	Sequence 9, Appl
641	6	0.8	1189	3	US-09-314-093-2	Sequence 2, Appl	814	1461	4	US-09-289-578-9	Sequence 9, Appl
642	6	0.8	1189	3	US-09-314-093-4	Sequence 4, Appl	815	1541	4	US-08-296-791-3	Sequence 3, Appl
643	6	0.8	1189	3	US-09-314-093-6	Sequence 6, Appl	816	1541	5	PCT-US95-10661A-3	Sequence 3, Appl
644	6	0.8	1189	3	US-09-314-093-8	Sequence 8, Appl	817	1545	4	US-08-296-791-4	Sequence 4, Appl
645	6	0.8	1189	3	US-09-314-093-10	Sequence 10, Appl	818	1545	5	PCT-US95-10661A-4	Sequence 4, Appl
646	6	0.8	1189	3	US-09-314-093-12	Sequence 12, Appl	819	1604	4	US-09-004-838-95	Sequence 95, Appl
647	6	0.8	1189	3	US-09-314-093-59	Sequence 59, Appl	820	1657	3	US-09-057-570-2	Sequence 2, Appl
648	6	0.8	1189	3	US-09-314-093-61	Sequence 61, Appl	821	1702	4	US-08-296-791-5	Sequence 5, Appl
649	6	0.8	1189	3	US-08-933-891-6	Sequence 6, Appl	822	1702	5	PCT-US95-10661A-5	Sequence 5, Appl
650	6	0.8	1189	4	US-09-250-848-2	Sequence 2, Appl	823	1805	3	US-09-057-570-7	Sequence 7, Appl
651	6	0.8	1189	4	US-09-250-848-4	Sequence 4, Appl	824	1805	4	US-09-004-838-92	Sequence 92, Appl
652	6	0.8	1189	4	US-09-250-848-6	Sequence 6, Appl	825	1817	4	US-09-004-838-125	Sequence 125, Appl
653	6	0.8	1189	4	US-09-250-848-8	Sequence 8, Appl	826	1843	4	US-09-413-814-50	Sequence 50, Appl
654	6	0.8	1189	4	US-09-250-848-10	Sequence 10, Appl	827	1848	4	US-08-296-791-6	Sequence 6, Appl
655	6	0.8	1189	4	US-09-250-848-12	Sequence 12, Appl	828	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
656	6	0.8	1189	4	US-09-176-320-8	Sequence 8, Appl	829	1938	4	US-09-514-302-2	Sequence 2, Appl
657	6	0.8	1189	4	US-09-251-885-2	Sequence 2, Appl	830	1940	2	US-08-644-271-30	Sequence 30, Appl

831	6	0.8	1940	4	US-09-077-955-34	Sequence 34, Appl	904	5	0.7	9	4	US-08-723-425A-337	Sequence 337, App
832	6	0.8	2182	2	US-08-487-826B-16	Sequence 16, Appl	905	5	0.7	9	4	US-08-723-425A-338	Sequence 338, App
833	6	0.8	2368	1	US-08-198-446B-15	Sequence 15, Appl	906	5	0.7	9	4	US-08-723-425A-339	Sequence 339, App
834	6	0.8	2368	2	US-08-870-693-15	Sequence 15, Appl	907	5	0.7	9	4	US-09-005-215-2	Sequence 2, Appl
835	6	0.8	2431	1	US-07-920-281C-2	Sequence 2, Appl	908	5	0.7	9	4	US-09-112-206-329	Sequence 329, App
836	6	0.8	2431	4	US-08-466-277-2	Sequence 2, Appl	909	5	0.7	9	4	US-09-112-206-330	Sequence 330, App
837	6	0.8	2439	3	US-09-335-409-7	Sequence 7, Appl	910	5	0.7	9	4	US-09-112-206-331	Sequence 331, App
838	6	0.8	2439	6	US-09-568-102-7	Sequence 7, Appl	911	5	0.7	9	4	US-09-112-206-332	Sequence 332, App
839	6	0.8	2439	6	US-09-567-969-7	Sequence 7, Appl	912	5	0.7	9	4	US-09-112-206-333	Sequence 333, App
840	6	0.8	2439	4	US-09-568-480-7	Sequence 7, Appl	913	5	0.7	9	4	US-09-112-206-335	Sequence 335, App
841	6	0.8	2439	4	US-09-568-486-7	Sequence 7, Appl	914	5	0.7	9	4	US-09-112-206-336	Sequence 336, App
842	6	0.8	2439	4	US-09-568-472-7	Sequence 7, Appl	915	5	0.7	9	4	US-09-112-206-337	Sequence 337, App
843	6	0.8	2439	4	US-09-567-899-7	Sequence 7, Appl	916	5	0.7	9	4	US-09-112-206-338	Sequence 338, App
844	6	0.8	2471	3	US-09-112-450-4	Sequence 4, Appl	917	5	0.7	9	4	US-09-112-206-339	Sequence 339, App
845	6	0.8	2471	4	US-09-419-291A-4	Sequence 4, Appl	918	5	0.7	9	4	US-08-772-282-29	Sequence 368, App
846	6	0.8	2584	4	US-08-936-135-4	Sequence 4, Appl	919	5	0.7	9	4	US-08-817-859-1868	Sequence 368, App
847	6	0.8	2588	3	US-08-936-135-2	Sequence 2, Appl	920	5	0.7	9	4	US-08-817-441-22	Sequence 22, Appl
848	6	0.8	2628	4	US-09-413-814-11	Sequence 11, Appl	921	5	0.7	10	1	US-08-460-463-4	Sequence 4, Appl
849	6	0.8	2756	1	US-08-375-709-11	Sequence 11, Appl	922	5	0.7	10	1	US-08-397-633A-87	Sequence 87, Appl
850	6	0.8	2756	1	US-08-752-929-11	Sequence 11, Appl	923	5	0.7	10	2	US-08-603-753D-24	Sequence 24, Appl
851	6	0.8	2756	4	US-09-090-793-7	Sequence 7, Appl	924	5	0.7	10	2	US-08-603-753D-25	Sequence 25, Appl
852	6	0.8	3038	1	US-08-450-332-2	Sequence 2, Appl	925	5	0.7	10	2	US-08-480-445A-4	Sequence 4, Appl
853	6	0.8	3038	2	US-08-637-640-2	Sequence 2, Appl	926	5	0.7	10	2	US-08-350-260A-338	Sequence 338, App
854	6	0.8	3038	4	US-09-004-406C-2	Sequence 2, Appl	927	5	0.7	10	2	US-08-460-446-4	Sequence 4, Appl
855	6	0.8	3739	3	US-09-320-878-2	Sequence 2, Appl	928	5	0.7	10	3	US-08-487-799-83	Sequence 83, Appl
856	6	0.8	3739	4	US-09-105-537-33	Sequence 33, Appl	929	5	0.7	10	3	US-08-159-339A-923	Sequence 93, App
857	6	0.8	4536	4	US-09-180-422B-27	Sequence 27, Appl	930	5	0.7	10	3	US-08-469-433B-4	Sequence 4, Appl
858	6	0.8	4928	4	US-09-036-987A-5	Sequence 5, Appl	931	5	0.7	10	4	US-09-099-753-25	Sequence 24, Appl
859	6	0.8	4928	4	US-09-370-700-5	Sequence 5, Appl	932	5	0.7	10	4	US-08-099-753-25	Sequence 25, Appl
860	6	0.8	10182	4	US-09-134-001C-3159	Sequence 3159, App	933	5	0.7	10	4	US-08-966-106-11	Sequence 21, Appl
861	6	0.8	11877	4	US-09-105-537-6	Sequence 6, Appl	934	5	0.7	10	4	US-08-966-106-22	Sequence 22, Appl
862	6	0.8	15281	2	US-08-471-119A-2	Sequence 2, Appl	935	5	0.7	10	4	US-08-075-541D-2	Sequence 2, Appl
863	5	0.7	5	2	US-08-288-059-5	Sequence 5, Appl	936	5	0.7	10	4	US-08-469-899A-4	Sequence 4, Appl
864	5	0.7	6	1	US-08-742-256-10	Sequence 10, Appl	937	5	0.7	10	4	US-09-194-062-1	Sequence 1, Appl
865	5	0.7	6	2	US-08-482-228-143	Sequence 143, App	938	5	0.7	10	4	US-09-187-859-3871	Sequence 3871, App
866	5	0.7	6	3	US-08-482-528-143	Sequence 143, App	939	5	0.7	10	4	US-08-197-484-60	Sequence 60, Appl
867	5	0.7	6	3	US-08-988-842-3	Sequence 3, Appl	940	5	0.7	10	4	US-08-197-484-139	Sequence 139, App
868	5	0.7	7	3	US-08-898-351-1	Sequence 1, Appl	941	5	0.7	10	5	PCT-US93-03406-4	Sequence 4, Appl
869	5	0.7	7	3	US-08-929-329-12	Sequence 12, Appl	942	5	0.7	10	5	PCT-US94-07916-4	Sequence 4, Appl
870	5	0.7	7	4	US-09-187-859-3862	Sequence 3862, App	943	5	0.7	10	5	PCT-US95-02121-60	Sequence 60, Appl
871	5	0.7	8	1	US-08-057-184-12	Sequence 12, Appl	944	5	0.7	10	5	PCT-US95-02121-139	Sequence 139, App
872	5	0.7	8	1	US-08-526-710-25	Sequence 25, Appl	945	5	0.7	10	5	PCT-US95-11127-17	Sequence 17, Appl
873	5	0.7	8	3	US-08-862-855-25	Sequence 25, Appl	946	5	0.7	11	1	US-07-699-468-1	Sequence 1, Appl
874	5	0.7	8	4	US-08-444-818-464	Sequence 464, App	947	5	0.7	11	1	US-07-699-468-2	Sequence 2, Appl
875	5	0.7	8	4	US-08-444-818-465	Sequence 465, App	948	5	0.7	11	1	US-07-699-468-3	Sequence 3, Appl
876	5	0.7	8	4	US-09-226-985-25	Sequence 25, Appl	949	5	0.7	11	1	US-07-699-468-4	Sequence 4, Appl
877	5	0.7	8	4	US-09-227-906-25	Sequence 25, Appl	950	5	0.7	11	1	US-07-699-468-5	Sequence 5, Appl
878	5	0.7	8	4	US-08-271-539-28	Sequence 28, Appl	951	5	0.7	11	1	US-08-136-743B-15	Sequence 15, Appl
879	5	0.7	8	4	US-08-271-539-29	Sequence 29, Appl	952	5	0.7	11	1	US-08-706-037-28	Sequence 28, Appl
880	5	0.7	8	4	US-09-187-859-3865	Sequence 3865, App	953	5	0.7	11	2	US-09-005-397-28	Sequence 28, Appl
881	5	0.7	8	5	PCT-US94-14106-49	Sequence 49, Appl	954	5	0.7	11	3	US-09-008-481A-11	Sequence 11, Appl
882	5	0.7	9	1	US-08-136-743B-5	Sequence 5, Appl	955	5	0.7	11	4	US-09-309-592-11	Sequence 11, Appl
883	5	0.7	9	2	US-08-146-028-329	Sequence 329, App	956	5	0.7	11	4	US-09-113-977C-16	Sequence 16, Appl
884	5	0.7	9	2	US-08-146-028-330	Sequence 330, App	957	5	0.7	11	4	US-09-187-859-3874	Sequence 3874, App
885	5	0.7	9	2	US-08-146-028-331	Sequence 331, App	958	5	0.7	11	4	US-09-187-859-3874	Sequence 3874, App
886	5	0.7	9	2	US-08-146-028-332	Sequence 332, App	959	5	0.7	11	4	US-09-588-417-5	Sequence 5, Appl
887	5	0.7	9	2	US-08-146-028-333	Sequence 333, App	960	5	0.7	11	6	5169773-1	Patent No. 5169773
888	5	0.7	9	2	US-08-146-028-335	Sequence 335, App	961	5	0.7	11	6	5202310-1	Patent No. 5202310
889	5	0.7	9	2	US-08-146-028-336	Sequence 336, App	962	5	0.7	11	6	5204096-15	Patent No. 5204096
890	5	0.7	9	2	US-08-146-028-337	Sequence 337, App	963	5	0.7	11	6	5204326-34	Patent No. 5204326
891	5	0.7	9	2	US-08-146-028-338	Sequence 338, App	964	5	0.7	12	2	US-08-406-330-48	Sequence 48, Appl
892	5	0.7	9	2	US-08-146-028-339	Sequence 339, App	965	5	0.7	12	2	US-08-406-330-69	Sequence 69, Appl
893	5	0.7	9	2	US-08-747-137-152	Sequence 152, App	966	5	0.7	12	2	US-08-556-597-48	Sequence 48, Appl
894	5	0.7	9	3	US-09-040-216-1	Sequence 1, Appl	967	5	0.7	12	2	US-08-556-597-69	Sequence 69, Appl
895	5	0.7	9	3	US-09-040-216-38	Sequence 38, Appl	968	5	0.7	12	2	US-08-726-306A-136	Sequence 136, App
896	5	0.7	9	3	US-08-159-339A-909	Sequence 909, App	969	5	0.7	12	2	US-08-129-030-43	Sequence 43, Appl
897	5	0.7	9	4	US-08-723-425A-329	Sequence 329, App	970	5	0.7	12	4	US-09-426-680-18	Sequence 18, Appl
898	5	0.7	9	4	US-08-723-425A-330	Sequence 330, App	971	5	0.7	12	4	US-09-055-075C-14	Sequence 14, Appl
899	5	0.7	9	4	US-08-723-425A-331	Sequence 331, App	972	5	0.7	12	6	5204326-35	Patent No. 5204326
900	5	0.7	9	4	US-08-723-425A-332	Sequence 332, App	973	5	0.7	13	1	US-07-991-199D-4	Sequence 4, Appl
901	5	0.7	9	4	US-08-723-425A-333	Sequence 333, App	974	5	0.7	13	1	US-08-503-062-2	Sequence 2, Appl
902	5	0.7	9	4	US-08-723-425A-335	Sequence 335, App	975	5	0.7	13	1	US-08-503-062-22	Sequence 22, Appl
903	5	0.7	9	4	US-08-723-425A-336	Sequence 336, App	976	5	0.7	13	1	US-08-191-338A-28	Sequence 28, Appl

977 5 0.7 13 4 US-09-196-293-6 Sequence 6, Appli
978 5 0.7 13 4 US-08-602-999A-88 Sequence 88, Appl
979 5 0.7 13 4 US-08-602-999A-89 Sequence 89, Appl
980 5 0.7 13 4 US-08-895-590-17 Sequence 17, Appl
981 5 0.7 13 4 US-08-209-603B-6 Sequence 6, Appli
982 5 0.7 13 4 US-08-278-865-88 Sequence 88, Appl
983 5 0.7 13 4 US-08-278-865-89 Sequence 89, Appl
984 5 0.7 13 4 US-08-278-865-90 Sequence 90, Appl
985 5 0.7 13 4 US-09-500-124-88 Sequence 88, Appl
986 5 0.7 13 4 US-09-500-124-89 Sequence 89, Appl
987 5 0.7 13 5 PCT-US93-12246-4 Sequence 4, Appli
988 5 0.7 13 5 PCT-US95-04121-10 Sequence 10, Appl
989 5 0.7 13 5 PCT-US96-11495-2 Sequence 2, Appli
990 5 0.7 13 5 PCT-US96-11495-22 Sequence 22, Appl
991 5 0.7 14 1 US-08-036-555B-126 Sequence 126, App
992 5 0.7 14 1 US-08-469-569-126 Sequence 126, App
993 5 0.7 14 1 US-08-249-322A-126 Sequence 126, App
994 5 0.7 14 1 US-08-642-255-103 Sequence 103, App
995 5 0.7 14 1 US-08-397-633A-5 Sequence 5, Appli
996 5 0.7 14 1 US-08-397-633A-10 Sequence 10, Appl
997 5 0.7 14 1 US-08-469-526A-126 Sequence 126, App
998 5 0.7 14 2 US-08-734-591A-126 Sequence 126, App
999 5 0.7 14 2 US-08-637-759B-250 Sequence 250, App
1000 5 0.7 14 2 US-08-469-660-126 Sequence 126, App

ALIGNMENTS

RESULT 1
US-08-656-177A-3
; Sequence 3, Application US/08656177A
; Patent No. 5882851
; GENERAL INFORMATION:
; APPLICANT: Koch, Birgit M.
; APPLICANT: Sibbesen, Ole
; APPLICANT: Halkier, Barbara Ann
; APPLICANT: Moller, Birger L.
; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5882851artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/656,177A
; FILING DATE: 08-AUG-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19783/A/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-656-177A-3

Query Match 1.0%; Score 7; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 14 AATVLAA 20
Db 10 AATVLAA 16
|||||
RESULT 2
US-09-256-797-3
; Sequence 3, Application US/09256797
; Patent No. 6133417
; GENERAL INFORMATION:
; APPLICANT: Koch, Birgit M.
; APPLICANT: Sibbesen, Ole
; APPLICANT: Halkier, Barbara Ann
; APPLICANT: Moller, Birger L.
; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6133417artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/256,797
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/656,177
; FILING DATE: 08-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: S-19783/A/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-09-256-797-3

Query Match 1.0%; Score 7; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
QY 14 AATVLAA 20
Db 10 AATVLAA 16
|||||
RESULT 3
US-08-817-811-97
; Sequence 97, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.

APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 75 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-97

Query Match 1.0%; Score 7; DB 4; Length 75;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 407 ARGJAL 413
Db 23 ARGJAL 29

RESULT 4
US-08-817-811-21
Sequence 21, Application US/08817811
Patent No. 6174528
GENERAL INFORMATION:
APPLICANT: Cooper, Juan A.
APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-21

Query Match 1.0%; Score 7; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 407 ARGJAL 413
Db 23 ARGJAL 29

RESULT 5
US-08-817-811-20
Sequence 20, Application US/08817811
Patent No. 6174528
GENERAL INFORMATION:
APPLICANT: Cooper, Juan A.
APPLICANT: Relf, Wendy A.
APPLICANT: Good, Michael F.
APPLICANT: Saul, Allan J.
TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
TITLE OF INVENTION: COMPRISING SAME
NUMBER OF SEQUENCES: 97
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 78 amino acids
TYPE: amino acid


```
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-817-811-20
;
; Query Match 1.0%; Score 7; DB 4; Length 78;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 407 ARGGLAL 413
Db 27 ARGGLAL 33

RESULT 6
US-08-817-811-96
; Sequence 96, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.
; APPLICANT: Relf, Wendy A.
; APPLICANT: Good, Michael F.
; APPLICANT: Saul, Allan J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,811
; FILING DATE: 14-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 96/11944
; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: FBRC:005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-817-811-96
;
; Query Match 1.0%; Score 7; DB 4; Length 78;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 407 ARGGLAL 413
Db 27 ARGGLAL 33

RESULT 7
US-08-817-811-19
; Sequence 19, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.
; APPLICANT: Relf, Wendy A.
; APPLICANT: Good, Michael F.
; APPLICANT: Saul, Allan J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,811
; FILING DATE: 14-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 96/11944
; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: FBRC:005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-817-811-96
;
; Query Match 1.0%; Score 7; DB 4; Length 78;
; Best Local Similarity 100.0%; Pred. No. 27;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 407 ARGGLAL 413
Db 27 ARGGLAL 33

RESULT 8
US-08-817-811-95
; Sequence 95, Application US/08817811
; Patent No. 6174528
; GENERAL INFORMATION:
; APPLICANT: Cooper, Juan A.
; APPLICANT: Relf, Wendy A.
; APPLICANT: Good, Michael F.
; APPLICANT: Saul, Allan J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDES AND VACCINES
; NUMBER OF SEQUENCES: 97
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,811
; FILING DATE: 14-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO 96/11944
; FILING DATE: 25-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: FBRC:005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 79 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; US-08-817-811-19
;
; Query Match 1.0%; Score 7; DB 4; Length 79;
; Best Local Similarity 100.0%; Pred. No. 28;
; Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 407 ARGGLAL 413
Db 30 ARGGLAL 36
```


SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,811
FILING DATE: 14-APR-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO 96/11944
FILING DATE: 25-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: FBRC:005
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 79 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-817-811-95

Query Match 1.0%; Score 7; DB 4; Length 79;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 407 ARGGLAL 413
|||
Db 30 ARGGLAL 36

RESULT 9
US-08-392-828C-39
Sequence 39, Application US/08392828C
Patent No. 5795962
GENERAL INFORMATION:
APPLICANT: IWANAGA, SADAAKI
APPLICANT: MUTA, TATSUSHI
APPLICANT: SEKI, NORIAKI
APPLICANT: ODA, TOSHIO
TITLE OF INVENTION: NOVEL POLYPEPTIDE AND DNA ENCODING
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 39
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
STREET: 53 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/392,828C
FILING DATE: 28-FEB-1995
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, PAULA A
REGISTRATION NUMBER: 32,503
REFERENCE/DOCKET NUMBER: FPN-033
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..127
OTHER INFORMATION: /note= "XLN A SEQUENCE (FIGURE 4)"
US-08-392-828C-39

Query Match 1.0%; Score 7; DB 1; Length 127;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 205 LNSDGSV 211
|||
Db 83 LNSDGSV 89

RESULT 10
US-09-330-945-39
Sequence 39, Application US/09330945
Patent No. 6077946
GENERAL INFORMATION:
APPLICANT: IWANAGA, SADAAKI
APPLICANT: MUTA, TATSUSHI
APPLICANT: SEKI, NORIAKI
APPLICANT: ODA, TOSHIO
TITLE OF INVENTION: DNA ENCODING HORSESHOE CRAB
TITLE OF INVENTION: AMEBOCYTE LYSATE FACTOR G SUBUNIT A
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/330,945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/119,995
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: FPN-032DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..127
OTHER INFORMATION: /note= "XLN A SEQUENCE (FIGURE 4)"
US-09-330-945-39

Query Match 1.0%; Score 7; DB 3; Length 127;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 LNSDGSV 211
Db 83 LNSDGSV 89

RESULT 11
5182262-8
; Patent No. 5182262
; APPLICANT: LETO, THOMAS
; TITLE OF INVENTION: CALMODULIN BINDING PEPTIDE DERIVATIVES
; OF NON-ERYTHROID ALPHA SPECTRIN
; NUMBER OF SEQUENCES: 15
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/318,172
; FILING DATE: 02-MAR-1989
; SEQ ID NO: 8
; LENGTH: 248
5182262-8

Query Match 1.0%; Score 7; DB 6; Length 248;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 HTVATEN 443
Db 139 HTVATEN 145

RESULT 12
US-07-941-651-2
; Sequence 2, Application US/07941651
; Patent No. 5332808
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5332808th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,651
; FILING DATE: 19920908
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-941-651-2

Query Match 1.0%; Score 7; DB 1; Length 278;

Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 558 KDKNAAA 564
Db 262 KDKNAAA 268

RESULT 13
US-08-279-996-2
; Sequence 2, Application US/08279996
; Patent No. 5552140
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5552140th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/279,996
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,651
; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 278 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-279-996-2

Query Match 1.0%; Score 7; DB 1; Length 278;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 558 KDKNAAA 564
Db 262 KDKNAAA 268

RESULT 14
US-08-746-797-2
; Sequence 2, Application US/08746797
; Patent No. 5759832
; GENERAL INFORMATION:
; APPLICANT: Gentry, Daniel
; APPLICANT: Lonsdale, John
; APPLICANT: Payne, David
; APPLICANT: Pearson, Stewart

TITLE OF INVENTION: NOVEL FabH
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,797
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5759832 Yet Assigned
FILING DATE: 23-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50573
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-746-797-2

Query Match 1.0%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 SLNSDGS 210
DB 178 SLNSDGS 184

RESULT 15
US-08-927-387-2
Sequence 2, Application US/08927387
Patent No. 5783432
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Lonsdale, John
APPLICANT: Payne, David
APPLICANT: Pearson, Stewart
TITLE OF INVENTION: NOVEL FabH
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,387
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/746,797
FILING DATE: 23-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50573
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
US-08-927-387-2

Query Match 1.0%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 SLNSDGS 210
DB 178 SLNSDGS 184

RESULT 16
US-08-918-058-2
Sequence 2, Application US/08918058
Patent No. 5885572
GENERAL INFORMATION:
APPLICANT: Gentry, Daniel
APPLICANT: Lonsdale, John
APPLICANT: Payne, David
APPLICANT: Pearson, Stewart
TITLE OF INVENTION: NOVEL FabH
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/918,058
FILING DATE: 25-AUG-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/746,797
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50573
TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 324 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; US-08-918-058-2

Query Match 1.0%; Score 7; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 204 SLNSDGS 210
Db 178 SLNSDGS 184

RESULT 17

US-08-921-887-52
; Sequence 52, Application US/08921887
; Patent No. 6030771

; GENERAL INFORMATION:

; APPLICANT: KHUYAKOV, YURI B.
; APPLICANT: FIELDS, HOWARD A.

; TITLE OF INVENTION: MOSAIC PROTEIN AND RESTRICTION
; ENDONUCLEASE ASSISTED LIGATION METHOD FOR MAKING THE SAME

; NUMBER OF SEQUENCES: 55

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: JONES & ASKEW, LLP

; STREET: 191 Peachtree Street, N.W., 37th Floor

; CITY: Atlanta

; STATE: GA

; COUNTRY: USA

; ZIP: 30303-1769

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/921,887

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: WARREN, WILLIAM L.

; REGISTRATION NUMBER: 36,714

; REFERENCE/DOCKET NUMBER: 03063-0380

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 404-818-3700

; TELEFAX: 404-818-3799

; INFORMATION FOR SEQ ID NO: 52:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 352 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: protein

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

; FRAGMENT TYPE: internal

; ORIGINAL SOURCE:

; ORGANISM: Hepatitis virus

US-08-921-887-52

Query Match 1.0%; Score 7; DB 3; Length 352;

Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 KQKALAF 368
Db 60 KQKALAF 66

RESULT 18

US-08-338-530A-3

; Sequence 3, Application US/08338530A

; Patent No. 5922327

; GENERAL INFORMATION:

; APPLICANT: CRABB, Brendan S.

; APPLICANT: STUDDERT, Michael J.

; TITLE OF INVENTION: EQUINE HERPESVIRUS GLYCOPROTEINS

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20007-5109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/338,530A

; FILING DATE: 25-JAN-1995

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/AU93/00253

; FILING DATE: 28-MAY-1993

; ATTORNEY/AGENT INFORMATION:

; NAME: BENT, Stephen A.

; REGISTRATION NUMBER: 29,768

; REFERENCE/DOCKET NUMBER: 40268/120/CSMB

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)672-5300

; TELEFAX: (202)672-5399

; TELEX: 904136

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 411 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-338-530A-3

Query Match 1.0%; Score 7; DB 2; Length 411;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 TVLAALS 22
Db 3 TVLAALS 9

RESULT 19

US-09-267-384-3

; Sequence 3, Application US/09267384

; Patent No. 6193983

; GENERAL INFORMATION:

; APPLICANT: CRABB, Brendan S.

; APPLICANT: STUDDERT, Michael J.

; TITLE OF INVENTION: EQUINE HERPESVIRUS GLYCOPROTEINS

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Foley & Lardner

; STREET: 3000 K Street, N.W., Suite 500

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/267,384
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/338,530
FILING DATE: 25-JAN-1995
APPLICATION NUMBER: WO PCT/AU93/00253
FILING DATE: 28-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 40268/120/CSMB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 411 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-267-384-3

Query Match 1.0%; Score 7; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 TVLAALS 22
Db 3 TVLAALS 9

RESULT 20
US-09-457-046B-64
Sequence 64, Application US/09457046B
Patent No. 6287835
GENERAL INFORMATION:
APPLICANT: Croteau, Rodney et al.
TITLE OF INVENTION: Transacylases of the Pacifitaxel Biosynthetic Pathway
FILE REFERENCE: 53679
CURRENT APPLICATION NUMBER: US/09/457,046B
CURRENT FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 74
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 64
LENGTH: 461
TYPE: PRT
ORGANISM: Arabidopsis thaliana
US-09-457-046B-64

Query Match 1.0%; Score 7; DB 4; Length 461;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 668 TFDAMAA 674
Db 277 TFDAMAA 283

RESULT 21
US-08-468-812-8
Sequence 8, Application US/08468812

Patent No. 5935836
GENERAL INFORMATION:
APPLICANT: Vehmaaper, Jari
APPLICANT: M nyl, Arja
APPLICANT: Fagerstr m, Richard
APPLICANT: Lantto, Raija
APPLICANT: Paloheimo, Marja
APPLICANT: Suominen, Pirkko
APPLICANT: Lahtinen, Tarja
APPLICANT: Kristo, Paula
TITLE OF INVENTION: Actinomadura Xylanase Sequences and Methods
TITLE OF INVENTION: of Use
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERN, KESSLER, GOLSTEIN & FOX
STREET: 1100 New York Ave., N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,812
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/332,412
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,001
FILING DATE: 29-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bugalsky, Larry B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 1050.0340002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2540
TELEFAX: 202-371-2600
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 491 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
POSITION IN GENOME:
CHROMOSOME/SEGMENT: M64551
US-08-468-812-8

Query Match 1.0%; Score 7; DB 2; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 LNSDGSV 211
Db 447 LNSDGSV 453

RESULT 22
US-08-590-563-8
Sequence 8, Application US/08590563
Patent No. 6300114
GENERAL INFORMATION:
APPLICANT: M nyl, Arja
APPLICANT: Vehmaaper, Jari
APPLICANT: Fagerstr m, Richard
APPLICANT: Lantto, Raija

; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lantinen, Tarja
; TITLE OF INVENTION: Production and Secretion of Proteins of
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W. Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,563
; FILING DATE: 26-JAN-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/468,812
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/332,412
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/282,001
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugaisky, Lawrence B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 1050.0340003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 491 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: M64551
; US-08-590-563-8

Query Match 1.0%; Score 7; DB 4; Length 491;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 205 LNSDGSV 211
Db 447 LNSDGSV 453

RESULT 23
US-09-319-989-6
; Sequence 6, Application US/09319989
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; TITLE OF INVENTION: METHODS
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989

; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 554
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: HAP4
; US-09-319-989-6

Query Match 1.0%; Score 7; DB 4; Length 554;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 498 YDLNDNN 504
Db 518 YDLNDNN 524

RESULT 24
US-09-134-001C-5569
; Sequence 5569, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5569
; LENGTH: 557
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-5569

Query Match 1.0%; Score 7; DB 4; Length 557;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 KNGDYSS 57
Db 44 KNGDYSS 50

RESULT 25
US-08-656-177A-2
; Sequence 2, Application US/08656177A
; Patent No. 5882851
; GENERAL INFORMATION:
; APPLICANT: Koch, Birgit M.
; APPLICANT: Sibbesen, Ole
; APPLICANT: Halkier, Barbara Ann
; APPLICANT: Moller, Birger L.
; TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5862851artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/656,177A
APPLICATION NUMBER: US/08/656,177A
FILING DATE: 08-AUG-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19783/A/PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sorghum bicolor
IMMEDIATE SOURCE:
CLONE: P-450-Tyr
US-08-656-177A-2

Query Match 1.0%; Score 7; DB 2; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVAA 20
|||||
Db 10 AATVAAA 16

RESULT 26
US-09-256-797-2
Sequence 2, Application US/09256797
Patent No. 6133417
GENERAL INFORMATION:
APPLICANT: Koch, Birgit M.
APPLICANT: Sibbesen, Ole
APPLICANT: Halkier, Barbara Ann
APPLICANT: Moller, Birger L.
TITLE OF INVENTION: Cytochrome P-450 Monooxygenases
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6133417artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/256,797
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/656,177
FILING DATE: 08-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19783/A/PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Sorghum bicolor
IMMEDIATE SOURCE:
CLONE: P-450-Tyr
US-09-256-797-2

Query Match 1.0%; Score 7; DB 4; Length 558;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 AATVAA 20
|||||
Db 10 AATVAAA 16

RESULT 27
US-09-564-805-240
Sequence 240, Application US/09564805
Patent No. 6333403
GENERAL INFORMATION:
APPLICANT: Tavtigian, Sean V.
APPLICANT: Teng, David H.F.
APPLICANT: Simard, Jacques
APPLICANT: Rommens, Johanna M.
TITLE OF INVENTION: Myriad Genetics, Inc.
TITLE OF INVENTION: Chromosome 17p-Linked Prostate Cancer Susceptibility
FILE REFERENCE: 2318-258
CURRENT APPLICATION NUMBER: US/09/564,805
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/107,468
PRIOR FILING DATE: 1998-11-06
PRIOR APPLICATION NUMBER: 09/434,382
PRIOR FILING DATE: 1999-11-05
NUMBER OF SEQ ID NOS: 240
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 240
LENGTH: 661
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-564-805-240

Query Match 1.0%; Score 7; DB 4; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 478 ESKTLHK 484
|||||
Db 27 ESKTLHK 33

RESULT 28
US-09-347-878-16
Sequence 16, Application US/09347878C
Patent No. 6376210
GENERAL INFORMATION:
APPLICANT: Yuan, Chong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
FILE REFERENCE: 25885-1651
CURRENT APPLICATION NUMBER: US/09/347,878C
CURRENT FILING DATE: 1999-07-06
NUMBER OF SEQ ID NOS: 75


```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-347-878-16

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 714;
Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 KIESGEG 479
DB 523 KIESGEG 529
|||||

RESULT 29
US-09-405-728-2
; Sequence 2, Application US/09405728
; Patent No. 6391316
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: US/09/405,728
; CURRENT FILING DATE: 1999-09-24
; EARLIER APPLICATION NUMBER: US 09/267,749
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Haemophilus somnus
US-09-405-728-2

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 971;
Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 AAGFGTH 192
DB 537 AAGFGTH 543
|||||

RESULT 30
US-08-026-138E-3
; Sequence 3, Application US/08026138E
; Patent No. 5502166
; GENERAL INFORMATION:
; APPLICANT: Masayoshi MISHINA
; TITLE OF INVENTION: NOVEL PROTEINS AND GENES CODING THE SAME
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nishiohata Residence 1-107
; STREET: 5214, Nishiohata-machi
; CITY: Niigata-shi
; STATE: Niigata-ken
; COUNTRY: JAPAN
; ZIP: 951
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS v.5
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/026,138E
; FILING DATE: 26-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 39563/1992

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 16
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-347-878-16

Query Match
Best Local Similarity 1.0%; Score 7; DB 4; Length 714;
Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 KIESGEG 479
DB 523 KIESGEG 529
|||||

RESULT 29
US-09-405-728-2
; Sequence 2, Application US/09405728
; Patent No. 6391316
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Rioux, Clement
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF HAEMOPHILUS SOMNUS
; FILE REFERENCE: 9000-0049.20
; CURRENT APPLICATION NUMBER: US/09/405,728
; CURRENT FILING DATE: 1999-09-24
; EARLIER APPLICATION NUMBER: US 09/267,749
; EARLIER FILING DATE: 1999-03-10
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Haemophilus somnus
US-09-405-728-2

Query Match
Best Local Similarity 1.0%; Score 7; DB 1; Length 1239;
Pred. No. 3.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 ADVSGSL 205
DB 907 ADVSGSL 913
|||||

RESULT 31
US-08-687-399-7
; Sequence 7, Application US/08687399
; Patent No. 5928381
; GENERAL INFORMATION:
; APPLICANT: Toft, Annette H.
; APPLICANT: Marcher, Dorte
; APPLICANT: Pedersen, Hanne H.
; APPLICANT: Nilsson, Thomas E.
; TITLE OF INVENTION: A Combined Desizing and Bleaching
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5928381o No. 5928381disk of No. 5928381th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,399
; FILING DATE:
; CLASSIFICATION: 008
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4127.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
```


TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1385 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-687-399-7

Query Match 1.0%; Score 7; DB 2; Length 1385;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 407 ARGGLAL 413
Db 1001 ARGGLAL 1007

RESULT 32
US-09-413-814-46
Sequence 46, Application US/09413814
Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or
TITLE OF INVENTION: heteropolypeptide compounds
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 46
LENGTH: 1544
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-46

Query Match 1.0%; Score 7; DB 4; Length 1544;
Best Local Similarity 100.0%; Pred. No. 3.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 212 RGRVMAQ 218
Db 673 RGRVMAQ 679

RESULT 33
US-08-462-467B-4
Sequence 4, Application US/08462467B
Patent No. 6210899
GENERAL INFORMATION:
APPLICANT: Rosenbaum, Jan S
TITLE OF INVENTION: The Use of a BMP Protein Receptor
TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road

CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,467B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Herko, Bart S.
REGISTRATION NUMBER: 32,572
REFERENCE/DOCKET NUMBER: 5474R
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 627-0633
TELEFAX: (513) 627-0260
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1618 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
US-08-462-467B-4

Query Match 1.0%; Score 7; DB 4; Length 1618;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 405 LYARGGL 411
Db 580 LYARGGL 586

RESULT 34
US-08-462-467B-2
Sequence 2, Application US/08462467B
Patent No. 6210899
GENERAL INFORMATION:
APPLICANT: Rosenbaum, Jan S
TITLE OF INVENTION: The Use of a BMP Protein Receptor
TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,467B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Herko, Bart S.
REGISTRATION NUMBER: 32,572
REFERENCE/DOCKET NUMBER: 5474R
TELECOMMUNICATION INFORMATION:
TELEPHONE: (513) 627-0633
TELEFAX: (513) 627-0260
INFORMATION FOR SEQ ID NO: 2:


```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2887 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
US-08-4652-4678-2

```

Query Match 1.0%; Score 7; DB 4; Length 2887;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels

RESULT 35
 US-08-462-467B-8
 ; Sequence 8, Application US/08462467B
 ; Patent No. 6210899
 ; GENERAL INFORMATION:
 ; - APPLICANT: Rosenbaum, Jan S
 ; TITLE OF INVENTION: The Use of a BMP Protein Receptor
 ; TITLE OF INVENTION: Complex for Screening Bone Metabolism Actives and Cells
 ; TITLE OF INVENTION: Co-Transfected With a Type II BMP Receptor and a Type I
 ; TITLE OF INVENTION: BMP Receptor
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: The Procter & Gamble Company
 ; STREET: 11810 East Miami River Road
 ; CITY: Ross
 ; STATE: OH
 ; COUNTRY: USA
 ; ZIP: 45061
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/462.467B

```
Query Match      1.0%; Score 7; DB 4; Length 2887;
Best Local Similarity 100.0%; Pred. No. 6.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels
```

RESULT 36
US-08-159-339A-836
; Sequence 836, Application US/08159339A
; Patent No. 6037135

GENERAL INFORMATION:
APPLICANT: Kubo, Ralph T.
APPLICANT: Grey, Howard M.
APPLICANT: Sette, Alessandro
APPLICANT: Celis, Esteban
TITLE OF INVENTION: HLA Binding peptides and Their
TITLE OF INVENTION: Uses
NUMBER OF SEQUENCES: 1254
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 836:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-836

```
Query Match          0.8%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

RESULT 37
US-08-159-339A-891
; Sequence 891, Application US/08159339A
; Patent No. 6037135
; GENERAL INFORMATION:
; APPLICANT: Kubo, Ralph T.
; APPLICANT: Grey, Howard M.
; APPLICANT: Sette, Alessandro
; APPLICANT: Celis, Esceban
; TITLE OF INVENTION: HLA Binding peptides and Their
; TITLE OF INVENTION: Uses
; NUMBER OF SEQUENCES: 1254
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco

```


STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/159,339A
FILING DATE: 29-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/926,666
FILING DATE: 07-AUG-1992
APPLICATION NUMBER: US 08/027,746
FILING DATE: 05-MAR-1993
APPLICATION NUMBER: US 08/103,396
FILING DATE: 06-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Ellen Lauver
REGISTRATION NUMBER: 32,762
REFERENCE/DOCKET NUMBER: 018623-005030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
TELEX:
INFORMATION FOR SEQ ID NO: 891:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-159-339A-891

Query Match 0.8%; Score 6; DB 3; Length 10;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 RGPSSL 159
|||||
Db 5 RGPSSL 10

RESULT 38
US-08-191-866D-64
Sequence 64, Application US/08191866D
Patent No. 5783195
GENERAL INFORMATION:
APPLICANT: Cochran, Mark D.
APPLICANT: Macdonald, Richard D.
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus S-IBR-052 And Uses Thereof
NUMBER OF SEQUENCES: 99
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/191,866D
FILING DATE: 4 February 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

REGISTRATION NUMBER: 28,678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-191-866D-64

Query Match 0.8%; Score 6; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 YKIESG 477
|||||
Db 2 YKIESG 7

RESULT 39
US-08-185-949B-64
Sequence 64, Application US/08185949B
Patent No. 5874279
GENERAL INFORMATION:
APPLICANT: Mark D. Cochran
APPLICANT: Richard D. Macdonald
TITLE OF INVENTION: Recombinant Infectious Bovine
TITLE OF INVENTION: Rhinotracheitis Virus
NUMBER OF SEQUENCES: 104
CORRESPONDENCE ADDRESS:
ADDRESSEE: John P. White
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,949B
FILING DATE: 03-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 678
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 278-0525
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-185-949B-64

Query Match 0.8%; Score 6; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 472 YKIESG 477
|||||
Db 2 YKIESG 7

RESULT 40
PCT-US93-07306-45

Sequence 45, Application PC/TUS9307306
GENERAL INFORMATION:
APPLICANT: MARGOLIS, Richard U.
APPLICANT: RAUCH, Uwe
APPLICANT: MARGOLIS, Renee K.
TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20004

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07306
FILING DATE: 03-AUG-1993

PRIOR APPLICATION DATA: US 07/922,911
APPLICATION NUMBER: US 07/922,911
FILING DATE: 03-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, Guy K.
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: Margolis=1A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 17 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-07306-45

Query Match 0.8%; Score 6; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 207 SDGSVR 212
Db 3 SDGSVR 8

RESULT 41
US-08-520-933-2
Sequence 2, Application US/08520933
Patent No. 5981194
GENERAL INFORMATION:
APPLICANT: Jefferies, Wilfred A.
APPLICANT: McGeer, Patrick L.
APPLICANT: Rothenberger, Sylvia
APPLICANT: Food, Michael R.
APPLICANT: Yamada, Tatsuo
APPLICANT: Kennard, Malcolm
TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
as Diagnostic and Therapeutic Agents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/520,933
FILING DATE: August 31, 1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Shona S. McDiarmid
REGISTRATION NUMBER: 38,798
REFERENCE/DOCKET NUMBER: 7685-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398
TELEX: 06-23115
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-520-933-2

Query Match 0.8%; Score 6; DB 2; Length 19;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 MRGPSS 158
Db 1 MRGPSS 6

RESULT 42
US-09-285-040-2
Sequence 2, Application US/09285040
Patent No. 6455494
GENERAL INFORMATION:
APPLICANT: Jefferies, Wilfred A.
APPLICANT: McGeer, Patrick L.
APPLICANT: Rothenberger, Sylvia
APPLICANT: Food, Michael R.
APPLICANT: Yamada, Tatsuo
APPLICANT: Kennard, Malcolm
TITLE OF INVENTION: Use of p97 and Iron Binding Proteins
as Diagnostic and Therapeutic Agents
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/285,040
FILING DATE: 01-APR-1999
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 7685-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: 416-364-7311
TELEFAX: 416-361-1398
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 19 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-285-040-2

Query Match 0.8%; Score 6; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 153 MRPGSG 158
Db 1 MRPGSG 6

RESULT 43
US-08-256-747C-59
Sequence 59, Application US/08256747C
Patent No. 6037448

GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: GILLAM, Shirley
APPLICANT: OU, Dawei
APPLICANT: TINGLE, Audrey
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE
NUMBER OF SEQUENCES: 78
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,747C
FILING DATE: 06-OCT-1994
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-370
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-256-747C-59
Query Match 0.8%; Score 6; DB 3; Length 23;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 SSSVFA 27
Db 2 SSSVFA 7

RESULT 44
US-08-834-130A-59
Sequence 59, Application US/08834130A
Patent No. 6180758

GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: GILLAM, Shirley
APPLICANT: OU, Dawei

APPLICANT: TINGLE, Audrey
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,130A
FILING DATE: 14-APR-1997
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-686 MIS:jib
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

US-08-834-130A-59
Query Match 0.8%; Score 6; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 SSSVFA 27
Db 2 SSSVFA 7

RESULT 45
US-08-726-306A-118
Sequence 118, Application US/08726306A
Patent No. 5958684

GENERAL INFORMATION:
APPLICANT: van Leeuwen, Frederik Willem
APPLICANT: Burbach, Johannes Peter Henri
APPLICANT: Grosveld, Franklin G.
TITLE OF INVENTION: DIAGNOSIS METHOD AND REAGENTS
NUMBER OF SEQUENCES: 189
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1 Financial Center
CITY: Boston
STATE: MA
COUNTRY: US

ZIP: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/726,306A
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 95/20080.4
FILING DATE: 02-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/009,832

;
; FILING DATE: 01-Jan-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 96,048-A (3255/00784)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 24 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-726-306A-118

Query Match 0.8%; Score 6; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 SPRPAE 228
|||||
DB 4 SPRPAE 9

Search completed: December 18, 2002, 07:00:10
Job time : 30.7742 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:38:19 ; Search time 49.0531 Seconds

(without alignments)
3032.756 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776

Sequence: 1 MGQFNSVFRINMTAATVLA.....NTFNIGSERWTANLRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3244	85.9	708	16 Q9JXZ0	Q9JXZ0 neisseria m
2	832	22.0	802	16 Q9HML3	Q9HML3 pseudomonas
3	821.5	21.8	815	2 Q51339	Q51339 pseudomonas
4	788	20.9	725	16 Q9JW26	Q9JW26 neisseria m
5	784	20.8	725	16 Q9JXU3	Q9JXU3 neisseria m
6	755	20.0	753	16 Q8ZFY6	Q8ZFY6 versinia pe
7	736.5	19.5	729	16 Q8X8H4	Q8X8H4 escherichia
8	721	19.1	724	16 Q8ZQ10	Q8ZQ10 salmonella
9	656	17.4	734	2 Q9X6A5	Q9X6A5 bordetella
10	512	13.6	863	16 Q8YV33	Q8YV33 anabaena sp
11	486.5	12.9	820	16 Q8YV33	Q8YV33 anabaena sp
12	486.5	12.9	885	16 Q8YV33	Q8YV33 anabaena sp
13	483.5	12.8	854	16 Q8YV33	Q8YV33 anabaena sp
14	481	12.7	858	16 Q8YV33	Q8YV33 anabaena sp
15	479	12.6	851	16 Q8YV33	Q8YV33 anabaena sp
16	475	12.6	863	16 Q8YV06	Q8YV06 anabaena sp

17	471.5	12.5	733	16 Q926C7	Q926C7 rhizobium m
18	467.5	12.4	867	16 Q8YTP0	Q8YTP0 anabaena sp
19	465	12.3	828	16 P72602	P72602 synecocyst
20	463	12.3	695	16 Q8XR06	Q8XR06 raietonia s
21	460.5	12.2	851	16 Q8YTX5	Q8YTX5 anabaena sp
22	460	12.2	760	16 Q8X7W7	Q8X7W7 escherichia
23	459	12.2	708	16 Q9HUX3	Q9HUX3 pseudomonas
24	458	12.1	726	2 Q9X7K9	Q9X7K9 rhizobium 1
25	458	12.1	747	2 Q9JUP3	Q9JUP3 rhizobium 1
26	450.5	11.9	747	16 Q8U189	Q8U189 agrobacteri
27	450	11.9	819	16 Q8UK13	Q8UK13 agrobacteri
28	449.5	11.9	872	16 Q8YTT7	Q8YTT7 anabaena sp
29	441.5	11.7	753	16 Q88590	Q88590 pseudomonas
30	436.5	11.6	732	2 Q88599	Q88599 pseudomonas
31	432.5	11.5	732	16 Q91422	Q91422 pseudomonas
32	432	11.4	755	2 Q9F0F9	Q9F0F9 campylobact
33	429	11.4	732	2 Q86424	Q86424 pantoea agg
34	424	11.2	858	16 Q8YU05	Q8YU05 anabaena sp
35	423	11.2	724	16 Q926C6	Q926C6 rhizobium m
36	421.5	11.2	696	16 Q82908	Q82908 salmonella
37	420.5	11.1	853	16 Q8YU26	Q8YU26 anabaena sp
38	416.5	11.0	853	16 P72609	P72609 synecocyst
39	415.5	11.0	857	16 Q8YV33	Q8YV33 anabaena sp
40	412	10.9	689	16 Q8XV08	Q8XV08 raietonia s
41	411	10.9	714	16 Q92YCL	Q92YCL rhizobium m
42	410.5	10.9	762	2 Q9E293	Q9E293 zymomonas m
43	410	10.9	801	16 Q8XV81	Q8XV81 raietonia s
44	407.5	10.8	635	16 Q9JTR6	Q9JTR6 neisseria m
45	402	10.6	863	16 P72599	P72599 synecocyst

ALIGNMENTS

RESULT 1

ID	Q9JXZ0	PRELIMINARY;	PRT;	708 AA.
AC	Q9JXZ0;			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	TonB-dependent receptor.			
GN	NMB1829.			
OS	Neisseria meningitidis (serogroup B).			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
OX	NCBI_TaxId=491;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=MC58 / SEROGROUP B;			
RX	MEDLINE=20175755; PubMed=10710307;			
RA	Tettelein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,			
RA	Risen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,			
RA	Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,			
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,			
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,			
RA	Cotton M.D., Ueberback T.R., Khouri H., Qin H., Vamathevan J.,			
RA	Gill J., Scariato V., Masigam V., Piazza M., Grandi G., Sun L.,			
RA	Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.,			
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain			
RT	MC58."			
RL	Science 287:1809-1815 (2000).			
RT	EMBL; AE002532; AAF42164.1; -.			
DR	TIGR; NMB1829; -.			
DR	InterPro; IPR002106; AATRNA_Ligase1.			
DR	InterPro; IPR001917; NHRfam2_2.			
DR	InterPro; IPR000531; TonB_BoxC.			
DR	Pfam; PF00593; TonB_BoxC; 1.			
DR	PROSITE; PS00599; AA_TRANSF CLASS 2; UNKNOWN 1.			
DR	PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN1_1.			
KW	Receptor; Complete Proteome			
SQ	SEQUENCE 708 AA; 77068 MW; F3519B024FBB4EE CRC64;			

Query Match

85.9%; Score 3244; DB 16; Length 708;

Best Local Similarity 86.8%; Pred. No. 1.4e-182;
Matches 627; Conservative 33; Mismatches 48; Indels 14; Gaps 3;

QY 1 MQQFMSVFRIINTAATVLAALSSVFAAQTABLETHVHIKQSYNAIVTEKNGDYSSFAV 60
DB 1 MQQFMSVFRIINTAATVLAALSSVFAAQTEGLETHVHIKQSYNAIVTEKNGDYSSFAA 60

QY 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRLVNSDDGRSSVYARG 120
DB 61 TVGTKIPASLREIPQSVSIITNQVKDRNVDTFDQLARKTPGLRLVNSDDGRSSVYARG 120

QY 121 EYSEYNIDGLPAQMOSINGTLNLFADFVEVMRGPGLFDSSGEMGGINLVLRKPTKA 180
DB 121 EYSEYNIDGLPAQMOSINGTLNLFADFVEVMRGPGLFDSSGEMGGINLVLRKPTKA 180

QY 181 FOGHAAAGFTGHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNRHETFYAAA 240
DB 181 FOGHAAAGFTGHQYKAEADVSGSLNSDGSVRGVMAQTVGASPRPAEKNNRHETFYAAA 240

QY 241 DWDINDPTVLGAGLYQQRHLPYNGLPADANNKLPSPHQHVFVGADWNKFKQNSHDVFA 300
DB 241 DWDINDPTVLGAGLYQQRHLPYNGLPADANNKLPSPHQHVFVGADWNKFKQNSHDVFA 300

QY 301 DLKHVFGNGGKGVGNYSRDRDADSNYAFAGSKLGKMTAGRPAGCNTADDKACAVGLGTE 360
DB 301 DLKHVFGNGGKGVGNYSRDRDADSNYAFAGSKLGKMTAGRPAGCNTADDKACAVGLGTE 360

QY 361 IKOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEQGRSTLYARGGLALNEFRSIP 420
DB 361 IKOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEQGRSTLYARGGLALNEFRSIP 420

QY 421 QYDLIANARKGVGRGSHYTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGDK 480
DB 421 QYDLIANARKGVGRGSHYTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGDK 480

QY 481 TLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVGYK 540
DB 481 TLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVGYK 540

QY 541 GSYMDRLNARYSFYRMKDKNAAAPLNPNKKTRVAALGKRVMEGVETEISGAVTPKQOI 600
DB 541 GSYMDRLNARYSFYRMKDKNAAAPLNPNKKTRVAALGKRVMEGVETEISGAVTPKQOI 600

QY 601 HAGYSLHSQIKTASNRDGDIFLLMPKHSANLWTTYQVTPBLTITGGGVNAMSGITSSAG 660
DB 601 HAGYSLHSQIKTASNRDGDIFLLMPKHSANLWTTYQVTPBLTITGGGVNAMSGITSSAG 660

QY 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGGANTENIPGSERTWTANLRY 720
DB 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGGANTENIPGSERTWTANLRY 720

QY 721 SF 722
DB 707 SF 708

RESULT 2
Q9HWL3 PRELIMINARY; PRT; 802 AA.

AC Q9HWL3
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE Probable TonB-dependent receptor.
GN PA4168.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
NCBI TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004833; AAG07555.1; -.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete Proteome.
SQ SEQUENCE 802 AA; 87430 MW; 379963AE03B47EAA CRC64;

Query Match 22.0%; Score 832; DB 16; Length 802;
Best Local Similarity 30.0%; Pred. No. 9.6e-41;
Matches 222; Conservative 137; Mismatches 309; Indels 72; Gaps 21;

QY 19 AALSSVFAAQTABLETHVHIKQ-QRSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQSV 77
DB 19 AALSSVFAAQTABLETHVHIKQ-QRSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQSV 77

QY 97 AATGAGIAPGSDVLDGALSVVATQDPATAAITEDSGSYTTRAMRTSTKLAMSIRETPQSV 156
DB 97 AATGAGIAPGSDVLDGALSVVATQDPATAAITEDSGSYTTRAMRTSTKLAMSIRETPQSV 156

QY 78 SIITNQVKDRNVDTFDQLARKTPGLRLVNSDDGRSSVYARGYSEYNIDGLPAQMOSI 137
DB 78 SIITNQVKDRNVDTFDQLARKTPGLRLVNSDDGRSSVYARGYSEYNIDGLPAQMOSI 137

QY 157 SVVTRQRMDDGMDLNDADVAGVTGLTVQQGPARKVYAAAGFVDNIMYDGLPTSTISY 216
DB 157 SVVTRQRMDDGMDLNDADVAGVTGLTVQQGPARKVYAAAGFVDNIMYDGLPTSTISY 216

QY 138 NG---TLNLPNLPADPDRVEMRGPGLFDSSGEMGGINLVLRKPTKAFQGHAAAGFTGHQ 194
DB 138 NG---TLNLPNLPADPDRVEMRGPGLFDSSGEMGGINLVLRKPTKAFQGHAAAGFTGHQ 194

QY 217 TDVISAADLAFDFRVEVVRGATGLMOGAGNPAAAINNVKKEPTQEFPRASLOGSVGTWDR 276
DB 217 TDVISAADLAFDFRVEVVRGATGLMOGAGNPAAAINNVKKEPTQEFPRASLOGSVGTWDR 276

QY 195 YKABADVSGSLNSDGSVRGR-VMAQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAG 253
DB 195 YKABADVSGSLNSDGSVRGR-VMAQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAG 253

QY 277 YRSEADVSGPLNSEGLRGAIVAYQKNSPDHVSSE-GVFYGIAEADLSPDTTFAIG 335
DB 277 YRSEADVSGPLNSEGLRGAIVAYQKNSPDHVSSE-GVFYGIAEADLSPDTTFAIG 335

QY 254 YLYQORHLAP-YNGLPADANNKLPSPHQHVFVGADWNKFKQNSHDVFPADLKHFGNGGYG 312
DB 254 YLYQORHLAP-YNGLPADANNKLPSPHQHVFVGADWNKFKQNSHDVFPADLKHFGNGGYG 312

QY 336 ASNQNDNRNDNWVGLPGLGRHLDLKRSSYYGADWSYMDTDTTHLFSDLTHRFANGWQM 395
DB 336 ASNQNDNRNDNWVGLPGLGRHLDLKRSSYYGADWSYMDTDTTHLFSDLTHRFANGWQM 395

QY 313 KVMRYSDRDADSNYAFAG-----SKLGKMTAGRPAGCNTADDKACAVGLGTEI 361
DB 313 KVMRYSDRDADSNYAFAG-----SKLGKMTAGRPAGCNTADDKACAVGLGTEI 361

QY 396 KLA-----ADKLWARINMLGLYNDCYSTTGCSMTQNFQDYSYTDH----- 439
DB 396 KLA-----ADKLWARINMLGLYNDCYSTTGCSMTQNFQDYSYTDH----- 439

QY 362 KOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEQGRSTLYARGGL---ALNEFRS 418
DB 362 KOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEQGRSTLYARGGL---ALNEFRS 418

QY 440 ----SYDAYANGPQLLGRHEHLVWGASVYRQERFDGHHGWSLFNKDGTPGMPDPTQW 494
DB 440 ----SYDAYANGPQLLGRHEHLVWGASVYRQERFDGHHGWSLFNKDGTPGMPDPTQW 494

QY 419 IPOVDLIANARKGVGRGSHYTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGE 478
DB 419 IPOVDLIANARKGVGRGSHYTVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKIESGE 478

QY 495 DPSSTLKPRLNTSLWGMK-----LDQEQKGAYLTRLNLADPLKVLGGRLDWDYKADADT 549
DB 495 DPSSTLKPRLNTSLWGMK-----LDQEQKGAYLTRLNLADPLKVLGGRLDWDYKADADT 549

QY 479 GKTLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVG 538
DB 479 GKTLHKASKTKFTYAGAVYDLNDNNSLYLSOLYTPQTNLDADGKLLKPRQGNQFEVG 538

QY 550 DS--YKVTN-NVTRYAGVIYDLNQTYSVYASYTDI FKPQSNFDAGGGLDITGKNYIEG 606
DB 550 DS--YKVTN-NVTRYAGVIYDLNQTYSVYASYTDI FKPQSNFDAGGGLDITGKNYIEG 606

QY 539 YGYSMDRLNARYSFYRMKDKNAA-----APLNPNKKTRYA--ALGKRVMEGVETE 587
DB 539 YGYSMDRLNARYSFYRMKDKNAA-----APLNPNKKTRYA--ALGKRVMEGVETE 587

QY 607 LKGEHFGGALNSQIALFOIDQENRATEDVGGPSCPFTSPT---SRYCSRASGKVRSGVD 663
DB 607 LKGEHFGGALNSQIALFOIDQENRATEDVGGPSCPFTSPT---SRYCSRASGKVRSGVD 663

QY 588 TEISGAVTPKQOIHAGYSLHSQIKTASNRDGD--IFLLMPKHSANLWTTYQVTPEL-- 643
DB 588 TEISGAVTPKQOIHAGYSLHSQIKTASNRDGD--IFLLMPKHSANLWTTYQVTPEL-- 643

QY 664 LELSGALSDDDQWQWAGYTYVDKAYKHDSNKANEGKPFDAKPRHLFKLATSYTLPELHK 723
DB 664 LELSGALSDDDQWQWAGYTYVDKAYKHDSNKANEGKPFDAKPRHLFKLATSYTLPELHK 723

QY 644 -TIGGVNAMSIGI-TSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGG 701
DB 644 -TIGGVNAMSIGI-TSSAGMHAGGYATFDMAAAYRFTPKLQINADNIFNRHYARVGG 701

QY 724 WRVGGDLATQSKTEDSSTGFGQGGYTVVNMGLYKYNRIDTRLNFNLLFKYIYSGIDF 783
DB 724 WRVGGDLATQSKTEDSSTGFGQGGYTVVNMGLYKYNRIDTRLNFNLLFKYIYSGIDF 783

QY 702 ANTFNIPGSERTWTANLRY 721
DB 702 ANTFNIPGSERTWTANLRY 721

QY 784 GN-LNY-GEPRNLMFTVKYS 801
DB 784 GN-LNY-GEPRNLMFTVKYS 801

RESULT 3

051339
ID 051339 PRELIMINARY; PRT; 815 AA.
AC 051339;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DN Ferrityoverdine receptor.
GN PPVA.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OK NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=93328663; PubMed=8335619;
RA Poole K., Neshat S., Krebs K., Heinrichs D.B.;
RT "Cloning and nucleotide sequence analysis of the ferrityoverdine
RT receptor gene fpyA of Pseudomonas aeruginosa.";
RT J. Bacteriol. 175:4597-4604(1993).
RN [2]
RP SEQUENCE OF 552-815 FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=95095977; PubMed=7798141;
RA Merriman T.R., Merriman M.E., Lamont I.L.;
RT "Nucleotide sequence of pvdD, a pyoverdine biosynthetic gene from
RT Pseudomonas aeruginosa: pvdD has similarity to peptidic synthetases.";
RT J. Bacteriol. 177:252-258(1995).
RN [3]
RP SEQUENCE OF 1-551 FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97075909; PubMed=8918232;
RA McMoran B.J., Merriman M.E., Rombel I.T., Lamont I.L.;
RT "Characterisation of the pvdB gene which is required for pyoverdine
RT synthesis in Pseudomonas aeruginosa.";
RT Gene 176:55-59(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; U07359; AAB60199.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KM Receptor.
FT CONFLICT 750 750 S -> R (IN REF. 1).
FT CONFLICT 716 716 Y -> F (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A11AE4A5A290F35 CRC64;
Query Match 21.8%; Score 821.5; DB 2; Length 815;
Best Local Similarity 29.0%; Pred. No. 4.1e-40;
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24;
QY 8 FRINMTAATVLAALSSVFAAQTADLETVAHIKQORSYNAIVTEKNGDYSSFAVVTGKIP 67
DB 108 PQGNATITISVAEADSSV-----DLAGMTITSNQL--GTTEDSGSYTPTGTTATATRLV 159
QY 68 ASARELPQSVITITNOQVAKRVNDTPQALRKTPGLVLSNDGSSVYARGYSEYNI 127
DB 160 LTPRETPQSVITITNOQVAKRVNDTPQALRKTPGLVLSNDGSSVYARGYSEYNI 127
QY 128 DGLPAGMOSI---NGTLPLVAFADRYEVMRSPGLPDSGEGGATVNLVRKPTAFQ 163
DB 220 DGLPSTARRVYAGNTLSMAIYDRVEVLKATGLTSGAGATNLIRKPTAFQ 163
QY 184 HAAAGFTHKQYAEADVSGSLNSDGSVGRVVAQTGASPPAEKNNRHETFYAADMD 243
DB 280 HVELGAGSMQNVNSLDSVGLPSTESGVNRGAVAAVQDKSPMDHYRKTSVYGLLEFD 339
QY 244 INPDYVLGAGYLYQQRHL--APYNG-LPA--DANNKLPGLFQHYFVGDANKKFKQNSHDVF 299
DB 340 LNDPTMLTVGADYQDNDPKSGWSGSEPLPDSQGNRNDVSRFNGAKWSSWEQYTRTVF 399

QY 300 ADLKHVFGNGYGVKGYRYSRDRDASNVAFAAGSKLGKTPAGRGCGNTADDKACAVGLGT 359
DB 400 ANLEHNPANGWGVKQL----DHKINGYHAPLAGIMG-DNPA-----PDNSAKIVAQXY 448
QY 360 EIQKALAPASYSRPPRLCANTANEVIG--ADYNFRFRSNEGRTTYLARG--GLALNEF 416
DB 449 TGETKSNSLDIYLTGPPQFGRHEHLVVGTSASFHW-----EGKSYWMLRNNDNTDDF 503
QY 417 RSIPOVDLANKARGVSHVATENLDFEFGYGSTFPAADGLSLGGRLGHYKIES 476
DB 504 -----INMGDICKPDMGIPSOYIIDKTRQLGSYMTAREVNTDNLDFLGGRVVDTRV-T 557
QY 477 GEGKTLHKASKTFTGYAGAVYDLDNNSLYLSLSQLYTPQTN--LDADGKLKPRQNG 534
DB 558 GLNPTIRESG--RFIYVGAVYDLDNTYSVYASYTIPMPQDSWYDSSNKLLEPDEGON 615
QY 535 FEVYGKSYMDPLNARVSEYRMDKNAAP-----LNPNNKTRVLAALG-KRVMEGVET 588
DB 616 YEIGIKGEYLDGRLNTSLAVFEIHEENRAEDALVNSKPTNPATYAYKGIKAKTKGYEA 675
QY 589 EISGAVTPKQIHAGYSYLHSQIKTASNSRDGIFILMPGHSANLMTTQO--VTBELTI 645
DB 676 EISGELAPGMQVAGYT--HKIIRDSGKK--VSTWEPQDLSLYTSYKFKALDKLTV 730
QY 646 GGGV-----NMSGITSSAGMHAGYATPDMAAAYFTPKLKLQINADNIFNR 693
DB 731 GGGARNGKSMQWVNNPPSRMEK---FQGEDYWLVDIARVYITDKLSASVAVNNVFDK 787
QY 694 HYARVAGANTENIPGSEERTANLRYSF 722
DB 788 TYTNTIGFYTSASY-GDPNIMFSTRWDF 815

RESULT 4
QY 050W26 PRELIMINARY; PRT; 725 AA.
AC 050W26;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DN Putative ferric siderophore receptor protein.
GN NMA0575.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OK NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z4491 / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holtroyd S.,
RA Jagers K., Leather S., Moulé S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis Z4491.";
RT Nature 404:502-506(2000).
DR EMBL; AL162753; CAB83866.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
KM Receptor; Complete proteome.
SQ SEQUENCE 725 AA; 80302 MW; DD6BF76A024EFPAB6 CRC64;
Query Match 20.9%; Score 789; DB 16; Length 725;
Best Local Similarity 29.9%; Pred. No. 3.2e-38;
Matches 232; Conservative 117; Mismatches 316; Indels 112; Gaps 23;
QY 5 MSVFRINMTAATVLAALSSVFAAQTADLETVAHIKQORSYNAIVTEKNGDYSS 57
DB 1 MTRFKYSLFPAALPYAAQADVSDDPKQESTELPITVTADRT-----ASSNDGYYTV 55


```

Qy 58 PAVTVGKIPASLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRVLSNDD--GRSS 114
Db 56 SGTHTPLGLPMTLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRVLSNDD--GRSS 115
Qy 115 VYARGVEYSEYNIDGLPAQMQSINGTLNPLFADFVEMRGPSGLFDSSGEMGGIVNLVR 174
Db 116 LPARSRIANYQINGIPVADALADTGNANTAYERVEVVRGVAGLLDGTGEPSTVNLVR 175
Qy 175 KRPTKAFQGHAAAGFGTHKQYKAEDVSGSLNSDGSVRGRVMAQ--TVGASPRPAEKNRHH 233
Db 176 KRPTKPLFEVRAEAGNRKHFGLDGADVSGSLNAEGLRGLVSTFGRGDSMRQRER--SRD 234
Qy 234 ERYFAAADWDINDPDTVLGAGLYQQRHLL---APYNGLPADANNKPLPSLPQHVFGADWKNF 290
Db 235 AELYGILEYDIAPOQTRVHAGMDYQOAKETADAPLSYAVVDSQYATAFGPKDNPATNMAN 294
Qy 291 FQWNSHDVFDLKHVFGNGGKVGWYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADD 350
Db 295 SHRALNLFAGIEHRFN-----QDWKLKAEYDYTRSR--FRQPYGVAGVLSIDH 341
Qy 351 KACAVGL---GTEIKOKALAFDASYSRPRFLGNTANEFVIGADYNFRSTNEQGRTTLYA 407
Db 342 NTAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIAGINGVKYAS--NKYGE----- 395
Qy 408 RGLALNFRSIPQVDLLIANARKVGRGYSHTVA-----TENLDEFGI-----YKST 454
Db 396 -----RST-----IPNAPINAYEFSRTGAYPOPASFAQTIPQYGTTRIGGYLATR 441
Qy 455 FHPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLDNNSLYLSLSQ 512
Db 442 FRAADNLSLLGGRYRTRTGSYDSRTQGMVYSANRFTPTGTGIVFDLTGNLSLGSYSS 501
Qy 513 LYTPTQNLADGKLLKPRQGNQFVGYKGSYMDRLNARVSYRMDKN--AAAPLNPNK 570
Db 502 LFVPSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRNLNAAVYRARKNNLATAAGRDPSG 561
Qy 571 KXTRAAALGKRWEGVETEISGAVTPKQIHAGYSYLHSQIKTASNRDDGIFLLMPKHS 630
Db 562 -NTYYRAANQAKTHGWEIEVGRITPEWQIOAGYSQSCTRDQDGSRLNPDSV-----PERS 616
Qy 631 ANLWTTYQVTPEL---TIGGGV-----NAMSGITSSAGMHAGGYATFD 670
Db 617 FKLFTAYHFAPAPSGWTIGAGVRQWSETHDTPATLRIPNPAKARAADNSRQKAYAVAD 676
Qy 671 AMAAYRFTPKLKLQINADNIFNRHYARVGGANTFNIP-----GSERTWTANLRYSF 722
Db 677 INARYRFPRAELSLNVDNLFNKHYRTQ-----PDRHSYGALRTVNAAFTVRF 724

RESULT 5
Q9JXU3 PRELIMINARY; PRT; 725 AA.
AC Q9JXU3;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TonB-dependent receptor.
GN NMB1882.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T.C., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;

```

```

RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RL MC58."; 287:1809-1815(2000).
DR EMBL; AE002538; AAF42216.1; -.
DR TIGR; NMB1882; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Receptor; Complete proteome.
SQ SEQUENCE 725 AA; 80478 MW; D3953D4485FD8FAF CRC64;

Query Match 20.8%; Score 784; DB 16; Length 725;
Best Local Similarity 29.4%; Pred. No. 5 Se-38;
Matches 228; Conservative 118; Mismatches 320; Indels 110; Gaps 21;

Qy 5 MSVFRINNTAATVL-----AALSSSVFAAQTADLETVHIKQRSYNAINVTEKNGDYSS 57
Db 1 MTRFKYLLFAALLPVYAQADVSDDPKQESTELPTITVTDRT-----ASSNDGYTV 55
Qy 58 PAVTVGKIPASLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRVLSNDD--GRSS 114
Db 56 SGTHTPLGLPMTLREIPOSVSIIITNOQVKDRNVDTFQ--LARKTPGLRVLSNDD--GRSS 115
Qy 115 VYARGVEYSEYNIDGLPAQMQSINGTLNPLFADFVEMRGPSGLFDSSGEMGGIVNLVR 174
Db 116 LPARSRIANYQINGIPVADALADTGNANTAYERVEVVRGVAGLLDGTGEPSTVNLVR 175
Qy 175 KRPTKAFQGHAAAGFGTHKQYKAEDVSGSLNSDGSVRGRVMAQTVGASPRPAEKNRHE 234
Db 176 KRLTRKPLFEVRAEAGNRKHFGLDADVSGSLNTEGTLRGLVSTFGRGDSWRRERSRDA 235
Qy 235 TFYAAADWDINDPDTVLGAGLYQQRHLL---APYNGLPADANNKPLPSLPQHVFGADWKNF 291
Db 236 ELYGILEYDIAPOQTRVHAGMDYQOAKETADAPLSYAVVDSQYATAFGPKDNPATNMAN 295
Qy 292 KQWNSHDVFDLKHVFGNGGKVGWYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDK 351
Db 296 RHRALNLFAGIEHRFN-----QDWKLKAEYDYTRSR--FRQPYGVAGVLSIDHN 342
Qy 352 ACVAVGL---GTEIKOKALAFDASYSRPRFLGNTANEFVIGADYNFRSTNEQGRTTLYAR 408
Db 343 TAATDLIPGYWHADPRTHSASVSLIGKYRLFGREHDLIAGINGVKYAS--NKYGE----- 395
Qy 409 GGLALNFRSIPQVDLLIANARKVGRGYSHTVA-----TENLDEFGI-----YKSTF 455
Db 396 -----RST-----IPNAPINAYEFSRTGAYPOPASFAQTIPQYGTTRIGGYLATRF 442
Qy 456 HPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLDNNSLYLSLSQ 513
Db 443 RAADNLSLLGGRYRTRTGSYDSRTQGMVYSANRFTPTGTGIVFDLTGNLSLGSYSS 502
Qy 514 YTPQNLADGKLLKPRQGNQFVGYKGSYMDRLNARVSYRMDKN--AAAPLNPNK 571
Db 503 FVPSQKDEHGSYLKPVTTGNLLEAGIKGEWLEGRNLNAAVYRARKNNLATAAGRDPSG- 561
Qy 572 KTRAAALGKRWEGVETEISGAVTPKQIHAGYSYLHSQIKTASNRDDGIFLLMPKHS 631
Db 562 NTYRAANQAKTHGWEIEVGRITPEWQIOAGYSQSCTRDQDGSRLNPDSV-----PERS 617
Qy 632 NLWTTYQVTPEL---TIGGGV-----NAMSGITSSAGMHAGGYATFDDA 671
Db 618 KLFYAYHFAPAPSGWTIGAGVRQWSETHDTPATLRIPNPAKARAADNSRQKAYAVADI 677
Qy 672 MAAYRFTPKLKLQINADNIFNRHYARVGGANTFNIP-----GSERTWTANLRYSF 722
Db 678 MARVRFNPRAELSLNDNLFNKHYRTQ-----PDRHSYGALRTVNAAFTVRF 724

RESULT 6
Q8ZFY6 PRELIMINARY; PRT; 753 AA.
ID Q8ZFY6
AC Q8ZFY6;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)

```



```

DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE Putative iron-siderophore receptor.
GN YP01537.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Kung'u K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414149; CAC90360.1; -.
DR InterPro; IPR000531; T0NB boxC.1.
DR Pfam; PF00593; T0NB boxC.1.
DR PROSITE; PS01156; T0NB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Hypothetical protein; Complete proteome.
SQ SEQUENCE 753 AA; 83838 MW; 739512F31D328166 CRC64;

Query Match 20.0%; Score 755; DB 16; Length 753;
Best Local Similarity 28.3%; Pred. No. 2, 9e-36;
Matches 223; Conservative 126; Mismatches 307; Indels 122; Gaps 25;

QY 9 RIMMT-----AATVLAALSSSVFAQAOTADLETYHIKQSRYSN-----AIV 48
DB 14 RFNCTPIPKPAWVLAANVSLGCAYATDENNSQKRENNPANTTITVASPLRHAGV 73
QY 49 TENKGVSSFAVTVGKIPSLREIPQSVSIITNQCYKDNVDFDQAKTGLRLSLN 108
DB 74 TBSGSGINTSMTATGALNISARETPQSVSVLTQRDRDNLMSVEAVNNTIGISYRQF 133
QY 109 DGRSSVYARGYESEYNIDGL-----PAQMSINGTLPLNFAFRVYVMGSPGLFDSG 164
DB 134 DSDRPFQTSKGMVANNVNRGVATFYDTRFNYGDNLT-DTMDRDLIVVGAAGLMAAPG 192
QY 165 EMGIVNLVKKRPPTKAFOGHAAAGFTGHOKYKADVSGSLNSDGYRGVMAQTVGASP 224
DB 193 NPSAVINLVKKRPPTQDFRGVSAGVSGWEKMTALDISGLNBSGSRGFTVAY----- 247
QY 225 RPAEKN---NRHE-----TFPAADWDINPTVYAGATLYQQRHL---APYNGIPLA-DANN 273
DB 248 --EDKNSFVRYQOSKNPFGLIEIDVTPNTLFTFG-ADTQRTLTIRGGMFGGLPLFNSAG 304
QY 274 KLPSLQHFVFGADVMKFKKNSHVDVADLGHYFGNGSGYGVGRYSRDR-----DSN 326
DB 305 GRNVYASATATSDMASAERTTQTLSSLOHNDN-GWNIKGFTFPNDLRQDVMMPG 363
QY 327 YAFAGSKLGKTPAGPGCMTADDKACAVAGLGTETIKOKALAPDASYSRPRLGNTANEFV 386
DB 364 YPDQNTNIGM-----RPGSLSLIDGA-----RQGNVYIQVNGQYSLGRQHQLG 408
QY 387 IGAADYRFRSTMEQRTTLYARGGALNFRSIFQV-----DLIANARK 430
DB 409 LGNNRRONINDNYLAT-----CNAARTCPDLGDFTPQMOYPKPWSDKRAYGSK 460
QY 431 GVRGSHVATATNLDDEFGYIGKSTFHPADGLSLIGGRLGHYKIESGEGTLHKAS-KTK 489
DB 461 G-----RSDSSQYVTVQLSLIDPLTLILGRLTTWETRGDNFGTPONARKYNE 509
QY 490 FTGAGAVYDLANDNSLYLSQLYTPQTNLDADGKLKROGNOFEVGYKSGSYMDRLN 549
DB 510 FVYSGLTVDINDLSDLYVTSYTEIFNDENRRDNNLTLLAVSGQNTAEAGKGVAFNSLD 569
QY 550 ARVSFRMDKQ-----AAAPLPNNKKTIRYALGKRVWEGVETEISGAVTPKQIHAGY 604

```

```

DB 570 YSLAVEIRONNMPVDDTAPRLPDNSQPYFAVDGKTK-RGFEAYSGKTEPMNISAGY 628
QY 605 SYLISOIKTASNSRDDGIFLMPKHSANLWTVQV---TPELTIGGVNAMS-----GITS 657
DB 629 TQYN--VKLPSSVTPQPVTVTRKYLKFTYTLTGQLSDLTLLGGGVNQSOIARLSS 686
QY 658 SAGMH---AGGYATPDMAAAYRFTPKLKIQINADNIFNRHYARVAGANTFNIPGSERTW 714
DB 687 PIGIQIGQSSFAIYSIMSRHYQFTPOLSTLVNDNLFNGHYIYQIQYNGY-LLGAPRV 745
QY 715 TALLRYSF 722
DB 746 DATRYTF 753

RESULT 7
QY 08X8H4 PRELIMINARY; PRT; 729 AA.
ID 08X8H4
AC 08X8H4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE Outer membrane receptor for ferric iron uptake.
GN FHUE OR Z1741 OR ECS1480.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=63334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RT Welch R.A., Blattner F.R.;
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohehuro E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005320; AAG55848.1; -.
DR EMBL; AP002555; BAB34903.1; -.
DR InterPro; IPR000531; T0NB boxC.
DR Pfam; PF00593; T0NB boxC.1.
DR PROSITE; PS00430; T0NB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; T0NB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 729 AA; 81231 MW; BBF5ADAI33F0595B CRC64;

Query Match 19.5%; Score 736.5; DB 16; Length 729;
Best Local Similarity 27.3%; Pred. No. 3, 4e-35;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;

QY 3 QFMSVRIMTATV-LAALSSSVFAQAOTADLETYHIKQSRYSNAIVTEKNDYSSFAVT 61
DB 11 QYQATKPSLLAGCIALALPLSAFAAPATE-ETVIVEGSAT---APDGDENDYSTSTS 66
QY 62 VGTKIPASLREIPQSVSIITNQCYKDNVDFDQAKTGPRLVSLNDGRSSVYARGYE 121
DB 67 AGTKQMTQRDIPQSVTVIYSQRMEDQQLTIGEVENLTLSKSDQADRALYYSRGFQ 126
QY 122 YSEYNIDGLPAQMS--INGTLPLNFAFRVYVMGSPGLFDSGEMGIVNLVKKRPT 178

```


Db 127 IDNYVDGIPYFESRWNLGALSDMALFERVEVRGATGLTGTGNPSAIAINVRKHAT 186
Qy 179 -KAFQGHAAAGFGTHQYKABADVSGSLNSGSRVGRVMAQTGVA---SPRPAEKNNRHE 234
Db 187 SREFKGVDAEYSGWKNRYVADLQSPLTEDGKIRARI---VGGYQNNDSWLDRYANSEK 242
Qy 235 TFYAA-ADWDINDPTVLGAGLYQQRHL--APYNGLPA-DANNKLPSPQHVFGADWNK 290
Db 243 TFFSGIVDADLGDLTTLTLAGYEQRIDVNSPTWGLPRWNTDGSNSYDRARSTAPDWAY 302
Qy 291 FKMNSHDVADLKHYFGNGGKVGMRYSRDDADSNY-----AFAGSKLGMKTP----- 339
Db 303 NDKELKVFMTLKQRFADTQWATLNATHSEVEFDSKMYVDAYVYNKADGMLVGPYSNYP 362
Qy 340 ----AGRPGCNTADDKACAVGLGTE-----IKQALAFDASYSRPFR--LGNATANEF-- 385
Db 363 GFDYVGGTGMNSGKRVKVDALDLFADGSYELFGRQHNLMFGGSYSKQNNRYFSSWANIPPD 422
Qy 386 VIGADYNFRSTNEQRTILYARGGLALNEFRSIPQVDLIANARKGVGYSHTVATENLD 445
Db 423 EIGSPYN-FNG-----NFPQTD-----WSPOSQAQDDTT 450
Qy 446 EF-GYIGKSTFHPADGLSLIGGRGLHYKIESGEGKTLHKASKTKFTGYAGAVDLDNDNN 504
Db 451 HMKSLYAAATRVTLADPLHLILGARYTNVRDT-----LTYSMKQHTTYPYAGLVFDINDNW 506
Qy 505 SLYLSLSQLYTPQTNLDADGKLLKPRQNGQFVGVKGSYMDRLNARVSFYRMKDKNA-- 563
Db 507 STYASVTSIFQPNDRDSSGKYLAIPITGNVYELGLKSDWNSRLTTLTAIFRIEQDNVAQ 566
Qy 564 ---APLNPNKKTRVAALGKRWEGVETEISGAVTPKQIHAGYS-YLHQSQIKTASNSRD 619
Db 567 STGTPIPSNGSETAYKAVDGTGSKGVFELNGAIDNNQLTGFATRYI-----AEDNEG 620
Qy 620 DGIFLLMPKHSANLWTTQ--VTPLETTGGVNMMSGITSSA-----GMHAGGYATFD 670
Db 621 NAINPNLRTTVMFTSRLPWPMLTGVGGVWQNRVYTDVTPYGTFRAEQGSYALVD 680
Qy 671 AMAAYRTPKLLQINADNIFNRHYARVGGANTFIPGSERTWTANLRYSP 722
Db 681 LFTRYQVTKNFSLOGNVNLPDKTYDNTVEGSIVY---GAPRNFSTITGYQF 729

RESULT 8
Q8ZQ10
ID Q8ZQ10 PRELIMINARY; PRT; 724 AA.
AC Q8ZQ10;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Outer membrane receptor for Fe(III)-coprogen, Fe(III)-ferrioxamine B
DE and Fe(III)-thodotulic acid uptake.
GN FHUE OR STM1204.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856 (2001).
RL EMBL; AE008752; AAL20133.1; -.
DR InterPro; IPR000531; TonB_boxC.
Pfam; PF00593; TonB_boxC; 1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 724 AA; 80893 MW; 94622BFE3714BFC2 CRC64;

Query Match 19.1%; Score 721; DB 16; Length 724;
Best Local Similarity 27.7%; Pred. No. 2.8e-34;
Matches 208; Conservative 137; Mismatches 321; Indels 86; Gaps 20;

Qy 13 TAATVLAALSSSV-----PAAQTADLETVHIKQSRSYNAIVTEKNGDYSSFAVTVGTKI 66
Db 15 TAAPSLLAMGMAFPAFAAE---DTVIVEGETTADAVNEEQ-DYSMKTAAAGTKM 69
Qy 67 PASIREITPQSVSIITNQOVKDRNVDTFDLARKTPLGLRVLSNDDGRSSVYARGVESEYN 126
Db 70 PMTQDIPQSVSIVSOORMEDQQLTGLGEVMTWTTLGISGQADSRIYSYSRGFEIDNYM 129
Qy 127 IDGLPQMOS---INGTLPNLFAPDRVEVMRPGSLFDDSGEMGGIVNLVRKPT-KAFQ 182
Db 130 VDGIPTYFESRWNLGDALTDTALYERVEVVRGANGMLTGTGNPSASINMIRKHAITSREFK 189
Qy 183 GHAAAGFTHQYKAEADVSGSLNSGSRVGRVMAQTGVAAPAEKNNRHEIFYAAADW 242
Db 190 GNVSTEYGSWKNQORYMDLOSPLTADGNVRGRIVAGYQNNDSWLDRYNSKAPFSGIVDA 249
Qy 243 DINPDTVLGAGLYQQRHL--APYNGLPA-DANNKLPSPQHVFGADM-----NKFK 292
Db 250 DLGTTNLSAGYEQKIDVNSPTWGLPRWNTDGSKNYSRSTAPDWAYNNKEINKFF 309
Qy 293 MNSHDVFPADLKHYFGNGGYGVK---GMYSRDDADSNIYAFAGSKLGMKTP-AGRPOCNT 347
Db 310 VTLKQRFASWQATLNATHTEVFKDSKMYIDALVDKETGLTVSPYCASYPVVGTCWNS 369
Qy 348 ADDKACAVGLGTE-----IKQALAFDASYSRPFRLCNTANEFVIGADYNRRSTNEQG 401
Db 370 GKRKVDALDLFADGAYELFGRQHNLMFGGSYSKQNNRYFSAWANVPDDIGNFSAFNGF 429
Qy 402 RTLYARGGLALNEFRSIPQVDLIANARKGVGYSHTVATENLDEFGIYKSTFHPADGL 461
Db 430 PTHWAPQNLAQDD-----TTHMK--SLYAAATRLSLADPL 462
Qy 462 SLIGGRGLHYKIESGEGKTLHKASKTKFTGYAGAVDLDNNSLYLSLSQLYTPQTNLD 521
Db 463 HLILGARYTNVRDT---LTYSMKQHTTYPYAGLIYDINDNWSAVASYTSIFQPNKRD 518
Qy 522 ADGKLLKPRQNGQFVGVKGSYMDRLNARVSFYRMKDKNA--APLNPNKKTRYAA 577
Db 519 KAGQYLAPITGNNYEAGLKSMDWNSRLTTLTSLVFRIBQNNVAQATTPIPGSNGEFAWKS 578
Qy 578 LGRKVMSEGVETEISGAVTPKQIHAGYSYLHQSQIKTASNSRDDGIFLLMPKHSANLWTT 637
Db 579 TDGTVSGVFEVNGALTDNQWTFGATRY-----VAEDNEGNANPNLRTSVKLFTRY 633
Qy 638 QV--TPLETTGGGVNAMSIGI---TSSAG---MHAGGYATFDAMAAYRFTPKLQINAD 688
Db 634 RLPAIPELTGVGGVWQNRVYTDVTPYGTFRAEQGSYALVDLFTRYQVTKNFSVQGNIN 693
Qy 689 NIFNRHYARVGGANTFIPGSERTWTANLRY 720
Db 694 NLFDKTYDNTIDGSIVYGAPRNV-SLTANYQF 724

RESULT 9
Q9X6A5
ID Q9X6A5 PRELIMINARY; PRT; 734 AA.
AC Q9X6A5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Ferric alcaligin siderophore receptor.
GN FAUA.
OS Bordetella pertussis.
OC Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;

OC Bordetella.
 NCBI_TaxID=520;
 (1)
 RN SEQUENCE FROM N.A.
 RC STRAIN=UT25;
 RX MEDLINE=99429842; PubMed=10498707;
 RA Brickman T.J., Armstrong S.K.;
 RT "Essential role of the iron-regulated outer membrane receptor PaaA in
 alcaligin siderophore-mediated iron uptake in Bordetella species.";
 RL J. Bacteriol. 181:5958-5966(1999).
 DR EMBL; AF135154; AAD26430.1; -;
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_boxC; 1.
 KW Receptor.
 SQ SEQUENCE 734 AA; 81568 MW; 117D8D40AA8B37F3 CRC64;
 Query Match 17.4%; Score 656; DB 2; Length 734;
 Best Local Similarity 27.6%; Pred. No. 1.9e-30;
 Matches 201; Conservative 113; Mismatches 331; Indels 84; Gaps 21;
 QY 17 VLAALSSSVFAAQ-----TADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTRKIPAS 69
 DB 24 LGLALFQHALQAEARTGNDIAQIPASVGRREISD--LREGTNAVYTEMSTATGLTLS 81
 QY 70 LREIPOSVSIITNOQVKNVDTFDQAKTPGLRLVLSNDGRSSVYARGEYSEYNIDG 129
 DB 82 PRETPOSVSVTRQQLDDGLTDLGAILATAPGISVTRSDSNRYSFSGARGFITDINFDQ 141
 QY 130 LRPQMOS---INGTLPRILFAFDREVMRPSGLPDSGEMGIYINLVKRPYAFQGHAA 186
 DB 142 LVSPILSQMNYGSDIMDMAIYDHEIVRGATGLMTSGNSAANFPRKAPLEFAATFN 201
 QY 187 AGEFTHKQYKAEADVSGSLNSDSVGRVMAQTGASPRAEKNNRHETVYAAADWDINP 246
 DB 202 ASGSMQYVAGDADISPIREDGRIRSRILVAAYSQGSYVHFIDTRKRITGYVVSADLTP 261
 QY 247 DTYLGAGYLYQQRHLADY--NGLP-----ADANNKLPRLPOHVFVAGADWNNFKRNSHD 297
 DB 262 DFTLITTSVEYQHNSNMGFGSGLPFLFYSDGSRTPFNRSVAN-----NAPMAQDTBAT 314
 QY 298 VFADLKHYPGNGGYGKVMYSDRDADSNYAFAGSKLGMTP--AGRPGCNTADDKACAV 355
 DB 315 YFADLTHRFNDWMLKRAYSHTDGRYLMKRVYRGYDPDRHTGIIAADPARSNYDG----- 369
 QY 356 GLGTEIKOKALAFDASYRPFRLGNTANEFVIGADYNRFRSTNEOGRTLLYARGGLALNE 415
 DB 370 ----NDRDRIHF--SLSAFPEAFGLRHEVALG-----WMSIDNHSIDQRYAMWGPRA--- 415
 QY 416 FRSIPOVDLIANARKG--VRGYSHTVATEN--LDFEFGYKSTFHPADGLSLIGGRL 469
 DB 416 ----PAIGSFPPWRRAHIQPSWADTLSPADDVRTKQTGAIVLGRFALAEPLHLIVGDRW 471
 QY 470 GHKIKESGE-GKTLHKASKTKFTGYAGAVYDLDNNSLYLSLSOLYTPQTNLADGKLLK 528
 DB 472 SDMKTKQMYFGSRRERYIKQKFTPYAGLTVDINDYTAVASYTEIFOPQANARLTSGGILP 531
 QY 529 PROGOFEVGYKSYMDRLNARVSFYRMKQKAA-----APLNNNKKTRTAAALGKR 581
 DB 532 PLSKSTELGLKAAIYLEGRLNTAALFQTRQDNLAQYIPSSSLPGFPNMQASR--AASGAK 590
 QY 582 VMEGVETEISGAVTPKQOIHAGYSYLSQIKTASNSRDGIFLLMPGRSANLWTTYYOV-- 639
 DB 591 V-EGIDLEHAGQILPDWNI--GASYTHFTTKDASG--NPIINTNHPRSLPKLYTTRVLP 644
 QY 640 -TEELITGGVNNAMSGITSSAGMHAG-----GYATFDMAAVRFPFKLKLQINADNIF 691
 DB 645 ALHRLTVGGGVQDSRYQAAASPRGVNEVODSYALVSLMARDFPKKLSATLVNNLRF 704
 QY 692 NRYAYARVG 700
 DB 705 DKTYIDQIG 713

RESULT 10
 Q8YV33
 ID Q8YV33 PRELIMINARY; PRT; 863 AA.
 AC Q8YV33;
 DT 01-MAR-2002 (T-REMBLrel. 20, Created)
 DT 01-MAR-2002 (T-REMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (T-REMBLrel. 21, Last annotation update)
 DE Ferrichrome-iron receptor.
 GN AL12148.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NCBI_TaxID=103690;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Matanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003588; BAB73847.1; -;
 DR InterPro: IPR002106; AAcRNA_LigaseII.
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam: PF00593; TonB_boxC; 1.
 DR PROSITE: PS00339; AA_TRNA_Ligase_II_2; UNKNOWN_1.
 KW Receptor; Complete proteome.
 SQ SEQUENCE 863 AA; 95445 MW; 40P9EC04FE462D34 CRC64;
 Query Match 13.6%; Score 512; DB 16; Length 863;
 Best Local Similarity 25.0%; Pred. No. 6.8e-22;
 Matches 182; Conservative 122; Mismatches 303; Indels 122; Gaps 28;
 QY 47 IYTEKNGDYSSFAVTVGTRKIPASLREIPOSVSIITNOQVKNVDTFDQAKTPGLRLV 106
 DB 204 VVTGQONGYRVQDATATKTDPLRDIPQISQVPREVLEDRVRSILAETVSG--VV 261
 QY 107 SNDDGRSS-----VYARGYE-----YSEYNIDGLPAQMOSINGTLPRILFAFDREVMR 154
 DB 262 DGADYNGSPAQDFTIRGFEQGSFRNGYRDVNSYGLTV-----GTL-----ERVEVLK 310
 QY 155 GPGSLPDSGEMGIYINLVKRPPTKAFQGHAAAGF--GTHKQYKAEADVSGSLNSGSVR 212
 DB 311 GPASVLFGAVERPGIINVTYKQPLS--EPYQUGFEVGNRAFPQPSIDBSGPLNADKTL 368
 QY 213 GRVMAQTGASPRPAEKNNRHETVYAAADWDI--NDPTVLGAGYLYQQRHLADYNGLPAD 270
 DB 369 YFENASYSQSDGFQDVTNTVLTIAFTIAMKLGDRDRLTYEYINPKGTFPEQYTSILSD 428
 QY 271 ANNKLP-----SLPQHVFGADWNNFKRNSHVDYFADLKHIFGNGYIGKVMKRSR--DAD 324
 DB 429 -NFTLPRSFYQAYPNNAAYVNDTQKL-----GX--TLSHKFSDWQIR 468
 QY 325 SNYAFAGSKLGMTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSR----- 374
 DB 469 NNTSVYTSK-----NAEYTLATGVNDQSIRQRPQODEPFDQNTFGQIDLL 515
 QY 375 -PPRLGNTANEFVIGADYNRFRSTNEOGRTLLYARGGLALNFRS.IPOVDLIANARKGV 433
 DB 516 GKENTGISISHQILIGDFHNIDT-----PARVQRYNPND--INPNYINP 561
 QY 434 GYSH--TYATENDLDFGTYKSTFHPADGLSLIGGRLGHYIKESGE--GKTLHKASK 487
 DB 562 SFDYGRSSSTERFQTYGYLYLOQITFLDNLKLLIGRFDWISGENTDVNTVGTQNPDS 621
 QY 488 TKFTYAGAVYDLDNNSLYLSLSOLYTPQTNLADGKLLKPRQOGFEGYKSYMDR 547
 DB 622 SASFPRIGLVYQPSKSVSLYTSQSFVEITGVNPPGEIIFEPFRGTQYENAGTADLEGR 661
 QY 548 LNAVSPYRMKDKNAAAPLNPNKKTRVAAL-----GKRVMEGVETEISGAVTPKQOIH 601


```
Db 582 LSATLAAYQITKSNLTP-DPDPER---AALDYLIQVGEQSRGIELDVAGEILPGWKAI 737
Qy 602 AGSYLHSHQIKTANSRDDGIFLL-MPKHSANLWTTQVTP-----LTIQGG---VNAMS 653
Db 738 ASYAYTNAEV-TEDNIDIPVGNRLVSPKQASLWTTTFEQNSDLKGLGFLGYVGTFRS 796
Qy 654 GITSSAGHAGGYATFDMAAAYRTPKLLQINADNINFRHYARVGGANTFNIPGSEBT 713
Db 797 G-DSANGFEIPDYLRTDAAYIR-RDGFKAGINIRNLFDTDYIRTSDDGRTFLRRGAPFT 854
Qy 714 WTANLRYSF 722
Db 855 IIGSISWEP 863

RESULT 11
Qy 889 PRELIMINARY; PRT; 820 AA.
AC Q8YTW7
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2596.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Ref. 8:205-213(2001).
DR EMBL; AP003590; BAB74295.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 820 AA; 91035 MW; BBD0B2486468C451 CRC64;

Query Match 12.9%; Score 486.5; DB 16; Length 820;
Best Local Similarity 24.6%; Pred. No. 2e-20;
Matches 185; Conservative 112; Mismatches 353; Indels 103; Gaps 24;

Qy 15 ATVLAALSSSVFAAQ-----TADLETVHIKQRS--YNAIVTEKNGDYSSFAVTVGTQIP 67
Db 126 AGLILAVESTTATQPPATPTSDAPPAPIAQDDPIELVVTGEODRVVPTASTATKTD 185
Qy 68 ASLREIPQSVSIITNQVKDRNVDTFDQLARTKTPGLRVLSNDDGRSSVYARGYSEYNI 127
Db 186 TPVRDIPQSIQVIPIQLEDDQKTTIQEVLQNVSGVKNQVGGTD---AGYRIRGFDQ 242
Qy 128 DGLPAQMSQING-----TLPNLFAPRVEVMRGPGLFSSGEMGIVNLVRKRPYKA 180
Db 243 DG-----NFRNGFNDTFVSLVDTANIDRIEVLKGPASVLFGQAPPGGIINVTQKPLRT 297
Qy 181 FQCHAAAGFGTHQYKARADVSGSLNSGVS--RGRVMAQTVGASPRAEKNNRHETFYA 238
Db 298 PYAAELNVGNVAFYRPSDFISGPLTDDGSLLYRLNVAQNSGSF---RDYNFLERVEVA 354
Qy 239 -AADNDINPDTVLGAGLYLQORHLAPYNGLPADANNKLPSPHQHVFVGADWNKFKQNSHD 297
Db 355 PVITWNISDRTSLTFDLEVDQNDYLFDRGIPSIGRPAIP-IPISFVGLP-HVYNDSTFR 412
Qy 298 VFADLKHVFGNGGYGVKVMRYSRDRDADSNYAFAGSKLGMKTKTAPRGONTADDKACAVGL 357
Db 413 IGYRLEHDFSKDWQRNAPFSFVSGSSGTAYAGGYDL-----IDDOFAPITV 459
Qy 358 G-TEIKQKALAFDASYSRPRFLGNTANEFVIGADYNRRFSTNEQGRITLYARGGLALNEF 416
```

```
Db 460 SRDEFTRDIYTLQTEVVVGKFTGSIHQPLIGVELRR---NTWKYTSFDVADPILLDIF 515
Qy 417 RSPQVDLLANARKGVGRYSHVTATENLDEFGIYKSTFHPADGLSLIGGRLGHY-KIE 475
Db 516 NPNYDVELPATPDESTFSY-----TTRRDITLGIYVQDQITFADNLKLLVGGRFDFORKE 570
Qy 476 SGBGKTLHKASKTKFTGYAGAVVYDLNDDNSLSLSQLYTPQT--NLDADGKLLKPERQGN 533
Db 571 EGSESTASSELSSAFSPRIGIVYQPIQALSLVASYQSQFKPDRFFGRSASNEPFKTRGT 630
Qy 534 QFEVGYKGSYMDRLNARVSFYRMKDKNAAA--PLNPNKKTKRYAALGKRVMEGVETEIS 591
Db 631 QYEVGIKAD-ISEKLSATLAAVEITKTNTVVTSDPNDNLNLS---VQVGEQSRGIELDIG 685
Qy 592 GAVTPKQWQHAGSYLHSHQIKTASNSRDDGIFL-----LMPKHSANLWTTQVTPELTIG 646
Db 686 GEIVPGWNIISYTY-----TDAITSKDNITPVGNRIDNVNPEHAASLWTSYEL----- 733
Qy 647 GGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKL-----KLOINADN 689
Db 734 -----QSGDLKGLGFLGLYVGDYADVNTSLLSYFRTDSAIYYKRDNWRALNFRN 788
Qy 690 IFNRHYARVGGANTFNIPGSEBTWTANLRYSF 722
Db 789 LFNETYETSQARNTI-YPGAPFTVIGSFQI 820

RESULT 12
Qy 889 PRELIMINARY; PRT; 885 AA.
AC Q8YXV8
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome iron receptor.
GN ALL1101.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Ref. 8:205-213(2001).
DR EMBL; AP003584; BAB73058.1; -
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 885 AA; 98141 MW; AF7F6BAA0A37FE08 CRC64;

Query Match 12.9%; Score 486.5; DB 16; Length 885;
Best Local Similarity 23.9%; Pred. No. 2.2e-20;
Matches 179; Conservative 117; Mismatches 343; Indels 109; Gaps 26;

Qy 7 VFRINMTAAVTVALSSSVFAAQADLETVHIKQRSYNAIVTEKNGDYSSFAVTVGTGI 66
Db 191 IFEVAAPTAPTAPQESQVQPEQPS-----ARSEEPIELVVTGEQGYKVEDATGTRT 244
Qy 67 PASLREIPQSVSIITNQVKDRNVDTFDQLARTKTPGLRVLSNDDGRSSVYARGYSEYNI 126
Db 245 NTPIRDTPFSIQVVPVEEVIKDDQVQRVTDALRAVPG--VISQDAPVSA-----FESFN 295
Qy 127 IDGLPAQMSQINGTLPNLFA-----FDRVEVMRGPGLFSSGEMGIVNLVRKRP 178
```


Db 296 FRGSSQSLRNGIRDETICTAGSGVANIERIEVLGRGAGLFGSGSGGTIVITKOPL 355
 QY 179 KAFQGHAAAGFTHKQYKAADVSGSLNSDGSVGRVMAQTVGASPREAEK-----N 230
 Db 356 SSPFEYIEGTGVSFDIYEGRVDLTGPIARNBD-----TLVYLTAS---ASLGSFIDFVD 408
 QY 231 NRHETFFAADM--DINPDYVLGAGLYQQRHLAPYNGLPADANNKPLSLPQHVFVGADW 288
 Db 409 NERYFISPVLTWLDKXTNLTLEAEYLSTKN--PNYNGLPF--LGTVPJNPEGEIPLSNTL 465
 QY 289 NK--FKNSHVDVFA---DLKHYFGNGGYGKVGKMSDDADSNYAFAGSKGMTPAGRP 343
 Db 466 NEBFDKNDROVYVLGNFEBRSENNQFRSLAARQETQNEAIPLLELL----- 516
 QY 344 GCNTADDKACAVGLGTEIK--QKALAPDASYSRPRFGTANEFVIGADYNRPRSTEGQ 401
 Db 517 -----EDNRTLVRAQSLSTRVNYVLTNTNVGDFPKTGSIAHKLLFGFDLLRDRDTR 571
 QY 402 RTTYIYAGGLALNEFRSIPQVDLIAMARKVGVSHVATNENLDFGIYKGSTHPADGL 461
 Db 572 QDPAIE-----PINLFNRYVGTQTIGSSAPSSNFISSNT-----DGLGIYLODQMTLADNF 621
 QY 462 SLIGGGRGLG--HYKIESGEGTLHAKSKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQT 519
 Db 622 KILGGFPIIISQKQEDNDGETINFLQDEAPSPRIGLYOPTKNIISLGSYSRSFTQYVG 681
 QY 520 LDADGKLLKPRQGNQFEVGYKSYMDRLNARVSPYMKQKNA--PLNPNKKTYYA 577
 Db 682 TSPDQRLFOBERGTQYVGVIGKDWLDRKLTTLAFYQITRNNVSNBPDNG-----FLIQ 737
 QY 578 LGRKMEGVETESGAVTPKQIHAGYSYLHSQIKTASNSRDGIFL--LMPKHSANIMTT 636
 Db 738 TGEORSOGIEIDLAGELIPGKIIIGYAY--SDAKVTQDQFEGNLLNNTKHAFLMTT 795
 QY 637 YQVTPELTIGGVNAGSISSAGMHAG-----YATDAAAYFTFKLK 682
 Db 796 Y---ELQSGN---LQGLGFLGLYYIGERQDLSNSFELPSYFRTDAMPYR--RUNFR 846
 QY 683 LQINADNIFNRHYARVAGANTFNIPGS 710
 Db 847 AALININLFTVEYETAFDALSV--VPGA 873

Db 847 AALININLFTVEYETAFDALSV--VPGA 873

RESULT 13
 Q8YUKO PRELIMINARY; PRT; 854 AA.
 ID 08YUKO: PRT; 854 AA.
 AC 08YUKO: PRT; 854 AA.
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Ferrichrome-iron receptor.
 GN ALR2311.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iitiguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakasaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003588; BAB7310.1; -
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC.1.
 KM Receptor; Complete proteome.
 SQ SEQUENCE 854 AA; 94535 MW; A379CEFC075552C7 CRC64;

Query Match 12.8%; Score 483.5; DB 16; Length 854;

Best Local Similarity 24.1%; Pred. No. 3.2e-20;
 Matches 183; Conservative 134; Mismatches 328; Indels 113; Gaps 29;

QY 18 LAALSSVFAQAQTADLETVH-----IKGQSYNALVTEKNGDYSFAVAVGKRIASL 70
 Db 157 LAPASTIPAEQPEQSTETPRPAQTAENNEPELVLVTGNQDQRYVESSTATKIDPPL 216
 QY 71 REIPQSVIITMQCKDRNVDPDLAKRTPLGLVLSNDGRSS--VYARGYSEYNID 128
 Db 217 RDI PASVQVLPKEIIGDQRYVRLNELANNVSGVQGGSGYGLSSSGYFINGF---ESGFE 273
 QY 129 GLPAQMGOSINGTLNPLA--PDREYVMRGP--SGLPDSSGEMGIVNLYKRPKTAFOGHAA 186
 Db 274 GLRNGFDFGFTSPRDANIERVEFLGPASVLYGSANNQGVNITTKKPLPDPSSRVG 333
 QY 187 AGFGTHQYKAEADVSGSLNSDGSVRGV--MAQTVGASPRPAENNNHETFYA--ADWD 243
 Db 334 MTIGSYDPRFTIDFTPLDDSKSVLYRLNVAENSGSPFDIEN---ESFFISPVTVN 390
 QY 244 INPDYVAGAGLYQQRHLAPYNGLPADANNKPLSLPQHVFVG--ADWKKFKNSHVDVADL 302
 Db 391 ISPKTSMTFEYEVQKNVYTFDRGLL--PGNTFPQIPISRLGEGFNNABEISVFTYNL 448
 QY 303 KHYFGNG---GYKVGMYSDRDADSNYAFAGSKL--GMKTPAGRPCNTADDKACA 354
 Db 449 EHQPSDMKFRQGFNVTSIGNTRIA--RNTNFSPEFLDEDDQTLPR---TSETSDQEN 504
 QY 355 VGLGTEIKQALAPDASYSRPRFGTANEFVIGADYNRPRSTEGQRTLYIYAGGLALN 414
 Db 505 ISLQTEVSGK-----FNTGSI RHNVLLGVELAKKFT-----Y 537
 QY 415 EFRSIP--QVDL--IANARKGV--RGVSHVATENLDEFGYKSTFHPADGLSIGG 467
 Db 538 DFEAPFASIDIRNPVYGAQPGFDRSFAGEYGDNL---AVTFQNLIEFTPNKLALAG 594
 QY 468 R---LGHYKIESGEGTLHAKSKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQ--TNLAD 523
 Db 595 RFWIDSFDRDPVSNVTNNVSESNFSPRVGIYQPTNSFLVASWTNSFNPOFGFRSRT 654
 QY 524 GKLLKPRQGNQFEVGYKSYMDRLNARVSPYMKQKNA--PLNPNKKTYYAALGRVM 583
 Db 655 GESFKPETSQFEVGIKQEFDFKRLSATLAVFDITKXNVLTTPDNDNFS--VQIGQKS 712
 QY 584 EGVETEISGAVTPKQIHAGYSYLHSQIKTAS-----NSRDDGIFLMPKHSANIMTTQ 638
 Db 713 RGLFLDIAGELIPGKIIAYIYIDSVSXNDLRLNDLSGV---PNSASLMTTYE 768
 QY 639 VTEPLTIGGVNAGSISSAGM-----HAGGYATFDMAAYFTPKIKLQ 684
 Db 769 FQK-----GSLAGLGFGLVYVDREATLPTIKIISYVRTASIFYR--RDNMRAA 819
 QY 685 INADNIFNRHYARVAGANTFNIPGSERTWTANLRYSF 722
 Db 820 INIKNLFDETYE---SQSFYLVPAAPLTVIGTISFEF 854

RESULT 14
 Q8YMK9 PRELIMINARY; PRT; 858 AA.
 ID 08YMK9: PRT; 858 AA.
 AC 08YMK9: PRT; 858 AA.
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 DE Ferrichrome iron receptor.
 GN ALA4924.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN 11
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iitiguchi M., Ishikawa A., Kawashima K., Kimura T.,

[illegible]

DT	01-JUN-2001 (TrEMBLrel. 17, Created)
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	TonB-dependent receptor.
GN	CC0028.
OS	Caulobacter crescentus.
OC	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC	Caulobacter.
OX	NCBI_TaxID=155892;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=ATCC 19089 / CB15;
RX	MEDLINE=21173498; PubMed=11259647;
RA	Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA	Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R. B.,
RA	Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA	DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA	Kolony J.F., Smit J., Craven M.B., Khouri H., Shetty J., Barry K.,
RA	Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.
RA	Salzberg S.L., Venter J.C. Shapiro L., Fraser C.M.; "
RT	"Complete genome sequence of <i>Caulobacter crescentus</i> ."
RL	Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141 (2001).
DR	EMBL; AE005677; AAK22016.1; -.
DR	TIGR; CC0028. -.
DR	InterPro; IPR000531; TonB_boxC.
DR	Pfam; PF00593; TonB_boxC; 1.
KW	Receptor; Complete proteome.
SW	SEQUENCE 851 AA; 90124 MW; B75B166237FCD592 CRC64;

Query Match	12.7%	Score 479;	DB 16;	Length 851;
Best Local Similarity	24.7%;	Pred. No. 5.8e-20;		
Matches 198;	Conservative 126;	Mismatches 341;	Indels 138;	Gaps
Qy	10	INMTAATVLAALSSVFAAQTA	DTLVTHTKQRSYNAIVTEKNGDYSSP-AVTV-----62	
Db	95	LSIAAVAGVAGLG-----LAQAA-----	VAGPDSADAATTANADRSVSSVTIDAKRVA144	
Qy	63	--GTKIPASLRIPOSVSIITHQVKDRNVDTFDOLAKT	PGLRVLNSDDGR---SSVY116	
Db	145	DPSSSKPTAPLVDTPKSVTVIPAKIIIEQTAASLADILRTSPG	ITFGAGEGQPLADRP204	
Qy	117	ARGVEYSEYN--IDGLPAQWOSINGTLPNLNFAFDR	VEVMRGPGLPDSSEMGGIIVLNR174	
Db	205	IRG-QASANNVFDG-----VRDGGQVREIFNLQEV	VEVVKGPDSAYGGKSGGGSLNLS259	
Qy	175	KRPTKAFQGHAAAGFTHKQYKAEADVSGSLNSDGS	VRGRVMAQTVGASP-RPAEKNRHH233	
Db	260	KSPKADSFARGSVGVGTDAVVRATADLNHALNESVAVRLNLLA	-TQGDTPGRKSVSFD-317	
Qy	234	ETFYAADWDINP-----DTVLGAGYLYQORHLP	AVNGIPLPADANNKLP-----SL278	
Db	318	-----WGVAPSLAIGLDGDTQLTASYVHLEG	QDTPDYGVPPLTKTTPRTASGILDV369	
Qy	279	PQHVYFGADNWKPKMNSHDV--FADLKHYFGNGG	YGVKGMWRYS-----DRDADSNA328	
Db	370	DRRSFYGVASRDYQTKSDTATFA-IDHRI	DETILNRQVVRYSKSLNDYIVTNPBGDGA428	
Qy	329	-FAGSKLGMKTPAGRPCNCTADDKACA	AVGLGTETI--KOKALAFDASYSRPRLGNATANEF385	
Db	429	QFVGQWMMKR-----GTKTRWNP	TETVAATDLHGKKTFLGLEHSDVGLGLESLREEN--481	
Qy	386	VIGADYNRFRSTNEQ-----GRTTLYARGG	UALNEFRSIPQVDL---IANARKVGRY435	
Db	482	-LNATYSTFTTSGAACPTGFTIATTL	SLASAGDGCTLVYKPNDKDAWTVGINRAPARNV540	
Qy	436	SHTVATENLDFEFGYKSTFHPADGSL	IGGGRLGHYKYESGEGKTLH-----483	
Db	541	AKTTA-----LYGPDTVKFEKVLN	LGLRHDRYESKGVQDVATQANGVFTSVTYTP592	
Qy	484	KASIKTKETGY-AGAVYDNLN	NNSLYLSLSLYTP-----QTNLDADGK-----LLKP529	
Db	593	RSGSWAFNTYOVGLVYKPTPGSS	LYXSVYSTASTPPTGGISAGDONSMTATGTGNLATVOLPE652	

RESULT 15
Q9AC38
ID Q9AC38
AC Q9AC38

QY	530	RQGNQGEVYKSYMDRLNARISFRMRDKNAALPLNPNKTRTAAALGKRWMEGYETE	569
Db	653	EDSESEFAQAKNAVPHDITLALSALEFQTSRKAQIOID-----AFYAAQGEVEVKEFEFG	708
QY	590	ISGAVTPKQIHAGSYLHSQ-IKTASNSRSDGIFLL-HPKSHANLMTTYQVPELITIG	647
Db	709	VSNGITPKQVQGGYITMSELYRGAYTSYVNGDPLANTPKHSISSFTTYIKVTRKIALGG	768
QY	648	GV-----NANSGITSSAG-----MHAGYATFDMAAAYRFTPKLQIOMNDINFRHYAR	698
Db	769	GAYHVSKEFGNGQGGAGGASRIYAPAYRYDPAFMAWAISTGVDDLOLNTQNLDERITYAR	828
QY	699	VGGATFN-IPGSERTWNLAKY 720	
Db	829	TNGVHHADPAFGROALITINVKY 851	
RESULT 16			
QY	Q8YV06	PRELIMINARY; PRT; 863 AA.	
AC	Q8YV06;		
DT	01-MAR-2002 (TREMBLrel. 20, Created)		
DT	01-MAR-2002 (TREMBLrel. 20, last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)		
DE	Ferrichrome iron receptor.		
GN	AKR2175.		
OS	Anabaena sp. (strain PCC 7120).		
OC	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.		
OX	NCBI_TaxID=103690;		
RN	SEQUENCE FROM N.A.		
RP	MEDLINE=21595285; PubMed=11759840;		
RX	Kaneke T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,		
RA	Watanabe A., Iriuch M., Ishikawa A., Kawashina K., Kimura T.,		
RA	Kishida Y., Kohara M., Matsumoto M., Matsumo A., Muraki A.,		
RA	Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,		
RA	Yasuda M., Tabata S.;		
RT	"Complete genomic sequence of the filamentous nitrogen-fixing		
RT	cyanobacterium Anabaena sp. strain PCC 7120."		
RL	DNA Res. 8:205-213(2001).		
DR	EMBL; AP003568; BAB73874.1; -		
DR	InterPro; IPR000531; TOMB_boxC.		
DR	Pfam; PF00593; TOMB_boxC; 1.		
KW	Receptor; Complete Proteome.		
QY	SEQUENCE 863 AA; 95343 MW; 3EDA3D64447D2372 CRC64;		
Query Match 12.6%; Score 475; DB 16; Length 863;			
Best Local Similarity 25.5%; Pred. No. 1e-19;			
Matches 176; Conservativeness 120; Mismatches 301; Indels 94; Gaps 26;			
QY	47	IVTEGNGDYSSFAVTVGKIPASLEIPQSVSIINQGVKDNVDFDGLAKRTG-LRV 105	
Db	201	VVTGADSYRVPNMAATRTDTPLELDIPQSVVPPQVLDQVNTLALQNVSGVQT 260	
QY	106	LSNDGSSVYARGEY-----SEYNDGLPAPOMOSINGTLLENLFAFDEVEYMRGSGLF 160	
Db	261	ASNYQSFASFTIRGFNSFDQGNFTNRLGLYRFGS-QGT-NFNSIERIEIVLRGGSVL 317	
QY	161	DSSGEMGIVNLVLRKRPKAFQGHAAAGFGTHKQYKAEDVGSLSNDSGVRGRVMAQTV 220	
Db	318	FGSGNPGGITINIVTKQPLSEPFYSVEAALIGSYDFRYKADIDLSGPIIDSDSTALYR-----L 372	
QY	221	GASPRPAK-----NNNHEFTYAA--DWIDINPTVLGAGVLVYQGRHLAPVNGLPADANNLT 275	
Db	373	NASYEKADNPVDFNDRENFVAATSLSPALGENTTLTLTLDQYKNAVQGYNGVPA-VGTVL 431	
QY	276	PSL-----DQHVFEVADWANKFKANSHDVFADLKGHYFGNGGYGKVGMRYS---DRDASNYA 328	
Db	432	PNLNGRIPIRNRISIGQADSYSPETVIRVGVNLEHKESEBDLLNNAFYISHFNKTKTDITYA 491	
QY	329	FAGSLGKMTAGRGRCGCTADDKACAVGLGIEIKQYALFADASYSRPPFLGNTANEFYVG 388	
Db	492	NA-----LDDPORTLQGVDDADDROYOTYDLSITVVK-----FSTGSIKHQLLFG 537	

QY	389	ADVNRERSTNEQRTLLVARGALALNEPFSISPOVD-----LINARKGVNGYSHITVATENL	444
Db	538	VDLSRY-----DQSSYSEGRGTPTDLEFNPDRPERFEVLFA-----GNEATLLT-----	582
QY	445	DEFGIGKSTFPHADSLSIGGRGLGHYKTESGE--GKTLHKASKTFGTGAVGAVDLDND	502
Db	583	DSLGVYIQOQVYFAENFKLLIGGRPFIEFQTKTRLSTNEQSGSAFSPRGGIVYQPLP	642
QY	503	NNSLYLSIQLYTPOTNLADGKLLKPRQGNFQEVGKYSYMDRLNARVSFPRMKDKNA	562
Db	643	PISLYASYSRSEFTPTIGRASDGEQFEPGRGTQYKIGVKAD--INEXLSATLAFYDLTRSNV	701
QY	563	AA--PLNPNKKTRRYAALGRWMEGTEITSGAVTPKQIHLGYSYLSHQITANSRSD	620
Db	702	TYNDPANPEFS-----IQTEQNSRGLIELVAGEIILPGWIIINGAYATYADR--TQDNSLP	756
QY	621	GIPL-LMPHGSANLMTTYVTPPELFTGGGVNMSGTSSAGNHAG-----	665
Db	757	GKRLNNVPEHSLSLMTTY-----ELQKGN-----LQSLGFLGLPYLQDRGDLANSFNLP	808
QY	666	YATFDMAAYRFTPKLKLQINDADINFNRRHY	696
Db	809	YLRDPAIFYK-RDRFRALNIRNLFDLKYF	838
RESULT	17		
ID	Q926C7	PRELIMINARY; PRT; 733 AA.	
AC	Q926C7;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)		
DE	Purative ferrichrome-iron receptor precursor protein.		
GN	PHUA1 OR R02223 OR SMC01611.		
OS	Rhizobium meliloti (Sinorhizobium meliloti).		
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;		
OC	Rhizobiaceae, Sinorhizobium.		
OX	NCBI_taxid=382;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=1021;		
RX	MEDLINE=21396507; PubMed=11481430;		
RA	Capela D., Batloy-Hudler F., Gouzy J., Botte G., Ampe F., Batut J.,		
RA	Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,		
RA	Godfrie T., Goffeau A., Kahn D., Kiss E., Lelaune V., Masny D.,		
RA	Pohl T., Portetelle D., Puehler A., Purnelle B., Rampeger U.,		
RA	Renard C., Thebaud P., Vandebol M., Weidner S., Galibert F.,		
RT	'Analysis of the chromosome sequence of the legume symbiont		
RT	Sinorhizobium meliloti strain 1021.'		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).		
EMBL	AL591790; CAC46802.1; -		
DR	InterPro; IPR000531; TOMB boxc.		
DR	Pfam; PF00593; TOMB_boxc; 1.		
KW	Complete proteome.		
SQ	SEQUENCE 733 AA; 79611 MW; 36D5C049C9B3786E CRC64;		
Query Match	12.5%; Score 471.5; DB 16; Length 733;		
Best Local Similarity	24.6%; Pred. No. 1,3e-19;		
Matches 192; Conservative 114; Mismatches 338; Indels 135; Gaps 30;			
QY	12	MTAAIVTALASSVA-----AQTADLETVHKKGSNYAIVTEKKGDYSSPAVTVGTPLP	67
Db	14	IASGVALPPLMMSGIALAQEGNATQLEIRIVVEGGAASATGVPDGVAAKATATGSGTGA	73
QY	68	ASIREIPQSVSITTTQGVKDRN-VDFPOLARKTPGLRVLS-----NDQGRSSVAVRGYEV	122
Db	74	MLNLTIPQSVSVSGEELDDRAVANKVDEALRYTPG--VLSAPFTDPTDPTMFTYRGFDA	131
QY	123	SEYN--IDGLPAQKQSIINGTLPLNLFAPDRVAVMKGPSGLFDSSGEMGIVNLVKKRPTKA	180
Db	132	AQTGLFLEGLPLFSGFGNFGVDFPMLERVENYTLKGPASVLYGSGSNPGIINLISKRPDLE	191

Wed Dec 18 08:35:28 2002

```
QY 181 FQHAAAGFGTHKQYKAEADVSGLSNDSGVRGVRMAQTVGASPRFAEKNNRHETTYAAA 240
D 192 PLYYTEVGINSNGNAFTGFDNKLDDGTVRYRLTKGVAGGN-----YSDY 239
QY 241 DWDINPDTVLGAGLYQOORHLAPYNGLPADANNKLSLPQ-HV-----FVG-----ADWN 289
D 240 SEDLR-----GFILPQVTVAPDDATSLTVFGLLOSLDQVHVGVNGFLPVYGVTEADPFG 292
QY 290 KFKMNSHVDVADL-----KHVFGNGYKGVKMYRSDRSDADSNYAFAGSKLG 335
D 293 KIDRDVYSEPDIDEGSYTQOMLGFEKFDHFDNGWTTQONARVANLHKHKEPYTYGVYG 352
QY 336 MKTPAGPGCNTADKAC-AVGLGTEIKQALAPDASYSRPFRLGNTANBEFVIGADYNRF 394
D 353 -----GAPDGYLLNRIGFPAATSKVDSFSDNRTETDFDGLATHTHFLAGLDVKKY 404
QY 395 RSTNEQ---GRTTLVARGGLALNEFRSIPQ-VDLIANARKGVRYGSHVTATENLDERGIY 450
D 405 RLDHIQACCATPI-----SATNPVYGTPOGANFV-----YLDQIVTQ--QQIGLY 448
QY 451 GKSTFHPADGSLIGGRLGHYKIESGEG-KTLHKASKTKETGYAGAVYDLDNNNSLYLS 509
D 449 AQDQIRFGDGLVTLNGRYDYVDTKSDAAIGTSYESNDGAFSGRAGLAIEFDNGLTPYVS 508
QY 510 LSQLYTP-----QTNLDADGKLLKPRQGNQFVG--YKGSYMDRLNARVSFYR 556
D 509 AATFFNPLVGTGSDPSDPTN-PAKMVALEPEEGYQYAGVKYEPSPFDGLLTASVFQIT 567
QY 557 MKDKNAAP--LNPNNKKTRYAALGKRVMEGVETEISGAVTPKQIHHAGSYLHSHQIKT 613
D 568 KQNVSIAPGFFVN-----SOLGEVRSRGEVLEKINLTNNKIIISAFSYTDLEVTE 619
QY 614 ASNSRDDG-IFLLMPKHSANLWTTQVTPELTIGGVNAMSGITSSAGMHAGVATPDAM 672
D 620 DLNASLIGNTPVLIPETQASLMDY-----TVANG--TFEGVSLGAGVRYOGESWADAE 671
QY 673 -----AAYRFTPK-LKLIQINADNINRHHYARVGGANTFNPISGERTWTANL 718
D 672 NTKKVPATLVDAATRYEKNDWTASLUNVANLFDKEVACQGLQICGY-GESRTFTLKL 729

RESULT 18
Q8YTP0 PRELIMINARY; PRT; 867 AA.
ID Q8YTP0;
AC Q8YTP0;
DT 01-WAR-2002 (TREMBLrel. 20, Created)
DT 01-WAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALL2674.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003590; BAB74373.1;
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC;
KW Receptor; Complete proteome.
SQ SEQUENCE 867 AA; 95468 MW; 5D21556D8E781FPA CRC64;
```

Query Match 12.4%; Score 467.5; DB 16; Length 867;
Best Local Similarity 24.7%; Pred. No. 2.8e-19;
Matches 186; Conservative 127; Mismatches 264; Indels 175; Gaps 38;

```
QY 61 TVGTKIPASLREIPQSVSLITNQOVKDRNVDTFDOLARKTPGLRVLSNDDGSSVYA---- 117
D 201 SVGTTRDTPLIDVPAIQIPIQEVINKDQTRSLGEVLKNTS-----SASSGRTSQAPAL 255
QY 118 ----RGYEYSEYNIDGLPAQMOSINGTILPNLPAFQFVEMRGPGLFDSSGEMGGIVNLV 173
D 256 TPVIRGFESRNLRLNGLRDSDLSRFQSEIANV---ERVELKGPASVLFQGGDLGGVNLV 312
QY 174 KRKP--TKAFQ--GHAAGFGTHKQYKAEADVSGLSNDSGVRGVRMAQTVGASPRPAEK- 229
D 313 TKOPLNTPYYSIGYGVQFGLH---RPTIDFSGPLDKG-----VAYRLNAAQYTAESF 363
QY 230 ---NNRHETTYAADWDI--NPDTVLGAG--YL-YQORHLAPYNGLP----- 269
D 364 KDPENSESFPIAPVVKLIGNENTNLNTASTEYLKYSFETAP--DLPASGTVISNPNGRVS 421
QY 270 -DANNKLPSPQHVFGVAGDWKNFKMNSHVDVADLKHFGNGGYGKVMYSDR---DADS 325
D 422 RETNLGEPULSE-----SESLVTRL-----GY-QLDHLRLNDNWTIKSEF 459
QY 326 NYAPAGSKLGMKTPAGR---PGCNTADDKACAVGL---GTEIKQ-----KALAFDA 370
D 460 STAF-----LDVPESTVVLPIANSSTSN---GLNRDGRTLRRLVENPSSTSLTFNN 509
QY 371 SYSRPFRLGNANEFVIGADYNFRSTNEQRTTLYARGGLALNEFRSIPQVDLI--ANA 428
D 510 SLGKFKTGSIEHSLLPFGEV---TRETEDEL---DFQLRDIDIFNPVYR 555
QY 429 RKGVRGYSHTVATENLDE--FGIYKSTFHPADGLSLIGGRLG----HYKIESGEGKTL 482
D 556 PESVSSFAIPFGNTNTEKNSIGIYAQOQISLSKNILVLGRLDFVNQDYEDLLSEESF 615
QY 483 HKASKTKFTGYAGAVYDLDNNNSLYLSLSQLYTP-----QTNLDADGKLLKPRQG 532
D 616 ER-NDTVFSPRGVIVYKPSNLSIYASYSRSTFPVQGRVLDNTGTTTGVGEFEPERG 674
QY 533 NQFEGYKGSYMDRLNARVSFYRMKDNAAA-----PLNPNKKTRYAALGKRVMEGVE 587
D 675 TQYEVGLKANLGLRSLTTLAFYNLERTNVNAAQGLSEPLS-----QIQIGKQRSQIG 727
QY 588 TEISGAVTPKQIHHAGSYLHSHQIKTASNSR-DDGI FLLMPKHSANLWTTQVTPELTIG 646
D 728 LDVAGEILPGHNLTSAYATDSKITEDSRPEFQDQRLQNVPRNSFGLWSTY----ELQAG 783
QY 647 GGVNAMSGITSSAGMHAGG-----YATPDMAAAYRTPKLLKLIQINADNIFN 692
D 784 ---SLKGLGFLGVFTQGERQGLRLNTFTLPSYLRTDASIFYR-RDKFRAAINQLNLF 838
QY 693 RHYIARVGGANTF--NIPGSERTWTANLRYSF 722
D 839 ENYYE---GARDIVRVIPGAPFTLTGSVSFEF 867

RESULT 19
P72602 PRELIMINARY; PRT; 828 AA.
ID P72602;
AC P72602;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FHUA OR SLL1406.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sabamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M.,
```


RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 DR EMBL; D90899; BA16602.1; -
 DR InterPro; IPR000531; TONB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR Complete proteome.
 SQ SEQUENCE 828 AA; 91420 MW; 315D8491FA82FE3 CRC64;

Query Match 12.3%; Score 465; DB 16; Length 828;
 Best Local Similarity 23.1%; Pred. No. 3.7e-19;
 Matches 182; Conservative 119; Mismatches 303; Indels 184; Gaps 30;

QY 17 VLAATSSSVFAA---QTADLETVHIKQGRSYNAIVTEKNGDVSSF--ATVGTGKIPASLR 71
 DB 143 VLSLTSSSTAIAPENPESIEV-----ATQEGGSEASYFVPSASTATGLDTPLL 192
 QY 72 EIPQSVSIITNOQVDRNVDFDQARKTGRLVLSNDGR--SSVARGYSEYNI--- 127
 DB 193 DIPQIQVVPQGVLDNRVTELGRLAQTVPG---VSPAGRGTSVFGGFLRGFPVNN 249
 QY 128 ---DGLPQ--MOSINGTLPNLPAPRVEVMRPSGLFDSGEMGIVLVKRRPTKARQG 183
 DB 250 IFRDIPYQSIAPLNTT-----DIEQIEVLKGPSSIVFGAGRPGGSIMLSKKPLDEPY 304
 QY 184 HAAAGFGTHKQVKAADVSG-----SLNSDSVGRVMAQTVGASPRPA 227
 DB 305 NNAVSLGYNDRLDVDSGLPLPAIDTVNRYLVSYETSISFDFYGDLMVVSF--- 361
 QY 228 EKNNHETFYAADMINDPVLGAGYLYOQRHLAPYNGLPADANNKLPSPLOHVFVGAD 287
 DB 362 -----TLTNIGBDTKLNTYGGYTTNRRLDSEIPAP---NIADLPSSNFLEGR 407
 QY 288 WNKFRMNSHDVFAADLKHFVNGGSGYKVGKRSYSDRDADSNYAFASKLGMKTPAGPGCNT 347
 DB 408 FSKFQDQYLLIGYTEFHHFENELKLRHAMQY-----LAVAPRYA-----PLRDF 452
 QY 348 ADDKACAVGLGEIKOKALAPDASRPRGLGNTANEF-----VIGADYRFRSTN 398
 DB 453 FDEDT-----GELNRFYYGGGNTQRFYNAELIGFEYTGPKRVFGLRYRNDTEP 506
 QY 399 EOGRTTLVARGGLALNEFRSIPQVDLIANARKVGYSHT---VATE-----NLDEFGY 450
 DB 507 EFGSNTFA-----PINVENPV-----YNTPEPIAPFERRDDQVNRFAVY 547
 QY 451 GKSTFHPADGLSLIGGGLGHYKLESSEGGKTLAKASKT-----KFTGYAGAV 497
 DB 548 LQDQMDLFDNLKLLVGLRY-----DSATQNRSTQSIITDPREEFNQDNLTFRGII 599
 QY 498 YDLINDNNSLYLSLSQLYTPQ--TNLDADGKLKPRQGNQFEGYGSYMDRLNARVFFY 555
 DB 600 YQPIPTVLSYLSYTTSPSPASISNADGSTFDPQTGRGFEGVAVAD--ITTKLSYTFSAF 658
 QY 556 RMKDKNAAPAPNPNNKTRRYAALGKRVMEGYETESGAVTEPKMOIHAGSYLSHQSIXAS 615
 DB 659 DIRQONVET-IDPANLLETIOT-GEQTSRGVELYIGGEILFGMINIVTGISYLDARVSODN 716
 QY 616 NSRDGIFLLMPKHSANIMTYYQVTPPELLTIGGVNAMSGITSSAGMAG----- 664
 DB 717 TDVNTLSNVPNSQFSLMTYEI-----QSGNIGQSGFGLGFVVDQREGDL 764
 QY 665 -----GYATFDMAAARFTPKLQIINADNIFNRHYAARGANTNI---PQSEETW 714
 DB 765 DNTFVLPSYFETDAIFVR-BENNELQINIENTLFTOYLAE---SNDPLSVYPGAPPTV 820
 QY 715 TANLRYSF 722
 DB 821 VGKIGVTF 828
 RESULT 20

Q8XR06
 ID Q8XR06 PRELIMINARY; PRT; 695 AA.
 AC Q8XR06;
 DT 01-MAR-2002 (TRENBLREL. 20, Created)
 DT 01-MAR-2002 (TRENBLREL. 20, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Probable ferrisiderophore receptor protein.
 GN RSP1062 OR RS02469.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC plasmid megaplasmid.
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoubat M., Genin S., Artiguenave F., Genzy J., Mangelot S.,
 RA Ariat M., Billault A., Broctier P., Camus J.C., Caticolico L.,
 RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaopin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissenbach J., Boucher C.A.;
 RL "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RL Nature 415:497-502(2002).
 DR EMBL; AL646082; CAD18213.1; -
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR Plasmid; Complete proteome.
 SQ SEQUENCE 695 AA; 75590 MW; FD7CA6956A201EBE CRC64;

Query Match 12.3%; Score 463; DB 16; Length 695;
 Best Local Similarity 24.5%; Pred. No. 3.8e-19;
 Matches 185; Conservative 112; Mismatches 336; Indels 122; Gaps 30;

QY 22 SSSVFAAQT-----DLETVHIKQGRSYNAIVTEKNGDVSSFAVGTGKIPASLR 72
 DB 5 AQAAPAEBAACAGPTVEAGTLPATVIGQTRSRIS-----YTPVATTKIPAPLRE 56
 QY 73 IPOSVSIITNOQVDRNVDFDQARKTGRLVLSNDGRSSVARGYE-YSEYNIDGLP 131
 DB 57 IPOSVNVVPRAVIODGALSLNDTLRVNVPVGSASLGDARDQVITIRGSAINDQVYDGLR 116
 QY 132 AQMSINOTLPNLPAPRVEVMRPSGLFDSGEMGIVLVKRRPTKARPGCHAAAGCT 191
 DB 117 DDALYFR---DLSNVRITELKGPAAVLVGRGSSGGIINRVTKPLATPMAEYVLLGT 172
 QY 192 HKQYKAADVSGSLNSDSVGRVMAQTVGASPRPAEKN---RHETFYAADMINDP 247
 DB 173 EGQGRABEDNTSINDA-----VRALLTGAVEDSGGRNDYFLRQAISPSFLPNISPD 227
 QY 248 T--VLGAGYLYQQR---HLAPYNGLPADANNKLPSPLOHVFVGADWNKFRMNSHDVFA 301
 DB 228 TKLTLQGFYLDHTRIDQGVPSYGRPVN---VPIETRYGSANAAGQANNVETTKSYGT 283
 QY 302 LKHFFGN--GGYKGYKMSYSDRDADSNYAFASKLGMKTPACRPGCNTAADKACVIGCTE 360
 DB 284 LDHRLINDQMSFHSVVRNYEYALGNNYTVVSRYTGAVPVTVLGVNORNSDRGTWQNE 343
 QY 361 IKOKALAFDASYSRPF--RLGNTANEFVIGADYRFRSTNBQGRITLYARGGLALNEFRS 418
 DB 344 LTOQAEFTFGIRHTLLYGLIELGYQDKSRVAA-----AGSFTYNNLPNP 386
 QY 419 IPQV--DLIANARKVGYSHTVATENLDEFGYKSTFHPADGLSLIGGRLGHYKI-- 474
 DB 387 VPVVLPTVPANATPSNYGLTNN-----ETVAMYAQLIFKSPQWYLAGLR--YEVLK 437
 QY 475 ESSEGGKT-----LHKASKTFGTGAGAVYDNLNDNNSLYLSLSQLYTP-----QTNLDA 522
 DB 438 QSRDLPKQODLSRTDK-PVSPRLGIVYHPVEALSLIYASYSRSPQWLADSFYYYNSSA 496
 QY 523 DGKLLKPRQGNQFEGYKGSYMDRLNARVS--FYRMKDKNAAPAPNPNNKTRRYAALG 579

497 ----LAPQSTTNYEIGAK---YDVSAASVSRAALPDMKQTN-----LTSVDPATQLAVPIG 545
580 KRMVEGVEITEISGAVTPKQWQHAGYSLHSQIKTASNRDDGIFL-----LMPKHSANL 633
546 TQTRGLELSFAGEVVKTVSWIASYAYLNG-VLTNPFDSKNGSVSGNQPSLTPRHSGSL 604
634 WTYQVTPETLTIGGVNA-----MSGTSSAGMHAGGYATPDMAAAYRFTPKLKLQIN 686
605 WVHDLDPNGFYVAGGLRAEGARFASQYVNT-----LPGYMTVDLGAGYR-SKHVDVTLN 658
687 ADMIFNRHYIYARV-GGANTFNIFGSERTWANTLRY 720
659 LQNLFRAYIVYSAHGAENYLNLPAGAPMALLGVRY 693
RESULT 21
ID Q8YTX5 PRELIMINARY; PRT; 851 AA.
AC Q8YTX5; 2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2588.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74287.1; --
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Receptor; Complete proteome.
KW SEQUENCE 851 AA; 93901 MW; 5F93DD9E558D291D CRC64;
SQ
Query Match 12.2%; Score 460.5; DB 16; Length 851;
Best Local Similarity 23.6%; Pred. No. 7.1e-19;
Matches 173; Conservative 120; Mismatches 316; Indels 123; Gaps 25;
QY 17 VLAALSSVFAAQTADLETVHIKQGSYNAIVTEKNGDYSSFAVVTGKIPASLREIPQS 76
164 VLSAAPSQTQATPTPEPEQPAATDQPIELVVTGQDQSVRYVNAATATKTDTPLRDIPQS 223
77 VSIITNQVDRNVDTFDQLARKTPGLRVLSNDGRS--SVYARGY-EYSEYNIIDGLPAQ 133
224 IQVTPROVEDQATQALNEALRNVSQVQ-SSNAGRTRDRFVIRGPDFDPSVIRDFG-KE 281
134 MQSINGTLNPLNFAFDRVEVMGPGSLFDSSGEMGIVNLVRKPTK--AFQGHAAAGFG 190
282 NTSVYRETANI--ERIEVLKGPASVLFQGLEPGGVINVTQKQREPFLLMGLEAGSYG 338
191 THQYKAEADVSGNSLSDGSVGRVMAQT-VGASPRPAEKNNRHETFYAAADWDINPDT- 248
339 F---FRPTVDNPSPLNDSTKLRLRVNAAVEISESRDFDKETSRFFIAPVLAWEIGDRTS 395
249 -VLGAGYLYQQRHAPYN-GLPADANNKLPSPQHFVFGADWNKFKMNSHVDVADLKHVF 306
396 LVFDLEYLKDSE--PFRDLGLVA-LGDRVAEIPFNRILGEPDPSKEVEDLRIGTRFEHGF 451
307 GNGGYGKGMVRRDSDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAV-----GLGTEI 361
452 NDWKLRSAPFRIVSTQTSFVTEPGSL-----DEATGLLSRDFGVDRT 495
362 KQKALAFDASYSRPFRLGNLTANEFVIGADYNRFRSTNEQGRTT-----LYARG 409

496 PDEYAFQTDLIGKFKTGAIEHELVIQDFNRQTNFVDDRSAPAPALDIENPVVLTARP 555
410 GLANNEPFIPOVDLIANARKGVRSYSHTVATENLDEFGIYKSTFFHPADGLSLIGGRL 469
556 NIESPPFRGL-----FAADNI---GIVIQQIKLAENLKLIGRY 593
470 GHYK---IESGEGKTLHKASKTKFTGYAGAVYDLDNNHSLYLSLSQLYTPQTNLDADGKL 526
594 DFTNQSFSLIVDGVKFFDIDSOAFSPRIGIVYQPTPLSLVASYSRSAQNFGIORDNSR 653
527 LKPRQGNQFVGKYSYMDRLNARVSFVRMKDKNAAP-----LNPNNKTKRYAALGKRV 582
654 IEPERGTOYEVGVRGEFLDGKLIASLAGYQITKTNVATPDADLNF-----IPUGEVR 707
583 MEGVETEISGAVTPKQWQHAGYSLHSQIKTASNRDDGIFL-LMPKHSANLWITYQVTP 641
708 SRGIEFDIAGELAKGNWIIASVAYTDAKI-TEDTNTDNEGNRLNRVPENSASITWITYEL-- 764
642 ELTIGGGVNAMSGITSSAGMHAG-----GYATFDAMAAYRFTPKLKL 683
765 -----QSCALQGLGNGVGLFFVGERQGDLSNFTVPGYTRTDALFYR-RDNWNI 813
684 QINADNIFNRHY 695
814 GLNFKNIFDVNY 825
RESULT 22
ID Q8X7W7 PRELIMINARY; PRT; 760 AA.
AC Q8X7W7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative outer membrane receptor for iron transport.
GN Z1026 OR ECS0883.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayaashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005261; AG55177.1; --
DR EMBL; AP002553; BAB34306.1; --
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 760 AA; 81956 MW; 60A294A347F7D502 CRC64;
Query Match 12.2%; Score 460; DB 16; Length 760;

Best Local Similarity 22.6%; Pred. No. 6,4e-19; Matches 168; Conservative 109; Mismatches 333; Indels 134; Gaps 23;

```
QY 65 KIPASLREIPOSVSIITITQVQVDRVDFPDQARLTKPGL-RILSNDDGRS-----SVYARG 119
Db 63 KFSRNVADTTMTYISQVVKDQATNLTDALKVPVGVGAFAEENNGNSTGDALYVRG 122
QY 120 YEYS- EYVIDGLPAQMGISNGLTNLPALFDRVEYVRGSPGLFDSGSGMGIVLVKRPPT 178
Db 123 ADTSNSIYIDGI---RDIGSVSRDTFMTQEVKIGKSPGTDYGRSAPLGSINMISKQPR 178
QY 179 KAFQHAAGAGTGHKQYAAEDVSSLSNDSGYSVRVWAQYVAGSPPRAEKNRHEITYA 238
Db 179 NDGSDASASIGSAMFRGLTDVNVQIDDTJAVRLNMGKTHDAGRDKVNER----- 232
QY 239 AADNDINDYLGAG-----YLQORHLAPYNGLP-----ADANKLPS 277
Db 233 ---YGVASIAFGIGTANRLYLVHTQHTPPDGIPTIGLPGYSASAGTATLTHSGK 289
QY 278 LPOHVFVAGADWNRKFNKSHDVFADLKHVFGNGYGVKGVKRYSDRDADSNY--AFAGSKL 334
Db 290 VDTNHFYGTDSDDSTDTDTATMRFEHDINDVTTIRNTTRMSRVKQD--YLTMTALMGAS 347
QY 335 GAKTPAGRPG---CNTDDACAVGLGTEIKOKALADASYSRPFLGNTANEVIGAD 390
Db 348 NITQPTSDVNSWTSWTSRTANTK-----DVSNKILITQMTLTSTFYASIGHDVSTGVE 399
QY 391 YNRPSTNEQGRITLYARGL---ALNEFRSIPQVDLIANARKGVRYSHVATENTLEF 447
Db 400 FTREIQTN-----YGVNPTVLPANVITHPDSSIHGGGLTFNGANANGQT-----DTF 446
QY 448 GIYKSTFHPADGLSLIGGRIGHYKIE-----SGEG-----KTLHK 484
Db 447 AIVAFDTQITRDFELNGIRLDNVHTEYDSATACGSGRGALITPAGVAKSPVTTTDT 506
QY 485 ASKTKFTGY-AGAVYDLNDNNSLYLSQLTTPQ-----TNLDADGKLKPRQG 532
Db 507 AKSGNLVMMKAGALYHLTENGVNYINAVSOQPPGANNFALAQSGSGNSANKTDFKPOKA 566
QY 533 NOFEVGYGYSYMDDLNARVSFYRMKDKAALPAPLNKKRYALIGRVMEGVETEISG 592
Db 567 NTSEIGTWQVLDKRLTLTALLFRDITENEY---EQNDGDTYSQYKKRVEGEIISVAG 622
QY 593 AVTPKMQIHAGYSYLSQIKTASNSRDDGIFLL--MPKHSANLMTTYOVTPELTIGGCVN 650
Db 623 NITPAMQVITGGTQOKATITKNGKDVADGSSSLPYTPBEHAFITMSQYQATDISGAGAR 682
QY 651 AMSGITSAGMHAG-----GYATPDMAAAYFTPKLQINADNIENHYYAR 698
Db 683 YI-----GSMHKGSDGAVGTPAFTEGVVADAKLGYRVNRMLDPQLNVVMLFDTDVVAS 736
QY 699 VGANTEFNI PGSERTW--TANLRY 720
Db 737 INKSGRYHPGEPRTFTLLTANMHP 760
```

RESULT 23

09HX3

ID 09HX3 PRELIMINARY; PRT; 708 AA.

AC 09HX3; 01-MAR-2001 (Tremblrel. 16, Created)

DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DE 01-MAR-2002 (Tremblrel. 20, Last annotation update)

DE Problem outer membrane protein.

GN PA4837.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxId=287;

RN [1]

RP SEQUENCE FROM N.A. STRAIN=ATCC 15692 / PAOI;

RX MEDLINE=20437337; PubMed=10984043;

```
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A.J., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AE004896; AAC08222.1; -.
DR HSRP; P06971.1; B5.
DR InterPro; IPR002106; AAcRNA_LigaseII.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 708 AA; 79063 MW; 08FBE0D639CBB CRC64;
```

Query Match

Best Local Similarity 12.2%; Score 459; DB 16; Length 708;

Matches 172; Conservative 135; Mismatches 329; Indels 122; Gaps 25;

```
QY 22 SSSVFAQT---ADLE--THIKGQRSTNAIVTEKNDYSPAVTGTKIPASIREIPQS 76
Db 16 SSPALMAETESPALEVLTVABEAREAGPV-----QGYRANRTASARTDRIEDIPQA 70
QY 77 VSITNQOVQRNVDTFQDLARKTPGLRVLSNDDGRS--SVYARGYEXSEYNIDGLPAQM 134
Db 71 ISVPRQVLDLDSARLERALDFAGVSRQNNFGSLTFEYNVAGFTTSEYRDGFSANR 130
QY 135 QSINGTLNLPALFDRVEYVRGSPGLFDSGSGMGIVLVKRPPTKAPQHAAGAGTGHKQ 194
Db 131 GYMNA--PDSATIRVEILKGPASSLYGRDGPGLVLTKKPQAEPRARLHAASGWSMR 188
QY 195 YKAEADYSGSLNSDGSVRGRMAQTGASPPRAEKNRHEITYA-----AADNDIN 245
Db 189 YRSTLDINTPLDEBGDLILYR-----NLAVEDSGFRDVAQGRLLVAPGISWQLD 239
QY 246 PDT-----VLGAGVYVQOORHLAPYNGLPADANLKLPSLPQHFVG--ADWNRKFNKSHD 297
Db 240 PDTSLVLEAEVYVRNRYQVDFRGTVAH-----NLGSLPRSPFGEPDQKLDNNET 291
QY 298 VFADLKHVFGNGYGVKGVKRYSDRDADSNVAFAGSKLGMKTPAGRPQGNATADYACAVGL 357
Db 292 LQATLRHFNQWGLRLASHYKHGLD--GYASENSSLA-----ADGYSL 334
QY 358 GTEIKOKALADASYR---PRLGNTANEVITGADYNFRSTNEQGRITLYARGGLA 412
Db 335 RREYRVRDFEWHDSITQLDLGLDHTGSIHQILMGLEYERYH--NDE-----LI 382
QY 413 LNEFRSIPQVDLIA--NARKGVRY-----SHVATENTLEFPIYKSTFHPADGL 461
Db 383 L---RSIPSRNPVALDIRPVGQPKPPFGRDNRH---EEVAMAMNLDDQIEFSKW 435
QY 462 SLIGGRLGHY-----KIESGE--GKTLHKASKTKFTGYAGAVYDLNDNNSLYLSQSL 513
Db 436 RGLGIVREFDRYQDMNATRLNNGRFRETSSQGTQRAATPRIGVLYQATPEVGLPANSKS 495
QY 514 YTPQNLDAADGKLKIPROGNOFEVGYKSYMDDLNARVSFYRMKDN--AAAPLNPNK 571
Db 496 FKPNQGTDMGAKAPDEBGRGEAGVXLDLDLGGITGLTAAFLHKKKRVVLTAADSPNPGYQ 555
QY 572 KTRAAALAGKRMEGVENEISGAVTPKMQIHAGYSYLSQIKTASNSRDDGIFILMPKQSA 631
Db 556 QT---AGEARSQGPDIQFSQQLTEQLRIGAVYVYIDAEVYKDNIAIGSRLLNVVPRHSG 611
QY 632 NLMTTYQVTEPLTIGGGVNA-----MSGITSAGMHAGGATATFDMAAAYFTPKLQ 684
Db 612 SLMGVEYERBEWGLGADAGAVNVVGRBAGDSGSGFELAPYTTVDLARYPLASNNATIG 671
QY 685 INADNIFNRHYAVGAGANTFNIPGSERTWTANLRYSP 722
Db 672 VVNVNLPDRRYER--SYNNVAVAPGEBPNLMTSLTANY 708
```



```

Db 316 DIRS---DPDYQKVKQOQIGYFHEHFNGLFRONLAFSHDLARLYGVSWTG- 371
Qy 337 KTPGRRCCTNADKACAVGLGTEIKOKALAPDASYSRPFGLTANFVIGADYRFRS 396
Db 372 -TVAHNRNASSIRDE-----MNVPQVNDQLEAKFPTGLAHMLFGLDY----- 413
Qy 397 TNEGRRTLYARGGLALNEFRSIPQVNDLIANARKVGR-----YSHTVATENDLDEFGTYG- 451
Db 414 -----TNLOSNNWGYGAGVNSAFDPD-IANPTYGVSGATPAYDPFVADAMRQVGYVAL 466
Qy 452 -----KSTFHPADGLSLIGGRLGHYKESGEGKTLHKASKTK 489
Db 467 DQIENAMRFPNLGGRQTFWVNVNTRDTYPSFGLN-----DSEVANKNA 508
Qy 490 FTGAGAVYDLDNNNSLYLSLSQLYTPQTNLDADGKLKPRQGNQFEVYKGSYM---DD 546
Db 509 FSWQAGALYLFEDNGIAFVSATSFDPVTNRASAGKILEPTEGQYELGVK--YQPGPTD 566
Qy 547 RLNARVSPYRHKDKNAAPLNNPKKTRYALGKRWMEGVETEISGANTPKQJHAGISY 606
Db 567 ILSAVA-YHIVKXKPVLVNP-LTLAYDSLGEVTGKIGILEKRAALADGIDLIIAATY 623
Qy 607 LHSQIKTASNSRDDGIFLLMKSHSANLWTTY---QVTP--ELTIGGVNAMSGITSSAGM 661
Db 624 NHESVTQGNDEGNTPAF--TPAHVASLMANYTFOETNPFNGLSVAGVRYVSEMWTDITAN 681
Qy 662 HAGGYATF--DAMAAFRFTPLKQINADNIFNRHYARVAGANTFNIP----- 708
Db 682 TSKNPSFFYDASAAYDF-----GAIDKNVEGLTAAFNIRINADQRDTVCNEGF 730
Qy 709 ---GSERTTANLRY 720
Db 731 CYLGQGRMTATLTKY 745

RESULT 26
O8U189 PRELIMINARY; PRT; 747 AA.
AC O8U189;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DE Ferrichrome iron receptor.
GN ATU0409 OR AGR_C_718.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kiteajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning U., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-D., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley P., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RL C58."
RL Science 294:2317-2323 (2001).
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Quirillo B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Tarchouk O., Bpp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;

```

```

RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328 (2001).
DR EMBL; AE009012; AAL41430.1; ALT_INIT.
DR EMBL; AE007978; AAK6224.1; -.
KW Receptor; Complete proteome.
SQ SEQUENCE 747 AA; 81287 MW; EEDC551426F61C9 CRC64;

Query Match 11.9%; Score 450.5; DB 16; Length 747;
Best Local Similarity 25.0%; Pred. No. 2.3e-18;
Matches 195; Conservative 104; Mismatches 358; Indels 123; Gaps 33;

Qy 2 GQNSVRINMTATVTAALSSSVPAAQTDLETVHKGSRVNAITYEKNQDYSSFAVT 61
Db 24 GMFLSRALGTAAVVLGMPVWG--HAQETTVLKQITVEGQGAENA--TGPVRYVANKSA 79
Qy 62 VGTIIPASLREIPOSVSILITNOQVXDN-VDTFQGLARKPFGLEAV--LSNDGSSVYAR 118
Db 80 TGSKTETETKAIPOSVSIVGRQEMDDRGAVTKIDEVLRTPGVTAEPPTDPTDWFYIR 139
Qy 119 GYEYSEYNI--DGLPAQMSINGTLPNLFAFDRVNRGSGFLPDSGEMGIVNLVYKR 176
Db 140 GFOATQGVFLDGLNLTSYGRGFGOMAYGLERVEVLKGPASVLYGKANPGIYQMSKR 199
Qy 177 ---PTKAFQ-----GHAAAGF-----GTHQYKAEADVSSGLN-SDGS--VYGR 214
Db 200 AQDTPVRETEIGINNFNAPFGFDLGDVDAEGVWK-YRVTGKVSGGDNYTDYSEDLRGF 258
Qy 215 VMAQTVGASPRPAKNNRHEFFVAAAWDINPVLGNGVLYQQRHLAPVNGLPADANAK 274
Db 259 IMPQ-ITFEP-DAQTSATLYGYFSALD-----QVHVENGLF-----PYVGTAVDA--P 302
Qy 275 LPSLPQHFVWG-ADWNKFKMNSHDVFPADLKHFGNGGKYGKMKRYS----- 320
Db 303 FGKLDRAKAFYCEPDIDNGRYQSVNGYDVSHPEPNKIKISOMARYGLYKHETPPYEGW 362
Qy 321 RDADSNYAFAGSKLGMKTPAGRPGCN--TADDKACAVGLGTEIKOKALAPDASYSRPFRL 378
Db 363 ANADAN-----GQPIIDPTNDYMLTRFGYDGLSKVDSFGVONRLEGGQFDT 408
Qy 379 GNTANEVIGADYRFRSTNBQGTITYANGGLALNEFRSIPQVNDLIANARKVGRG--- 434
Db 409 GAVNHSPLFGIDYXYRYLDVQVQACGSGNAIGAL-----KPYVGSITOG 450
Qy 435 ---YSHTVATENDLDEFGIYKSTFHPADGLSLIGGRLGHYKTESG-----GKTLHKASK 487
Db 451 TNEFYADNIVTQ--QQLGITAQDQLRFQDGLVTLNCRVYDTTELNNRLPAGVSRSSND 508
Qy 488 TKFTGAVAGVYDLDNNNSLYLSLSQLYTPQTNLDADGKLKPRQGNQFEVYKGSYMD 545
Db 509 DALSGRGLAVFPDNGLTTPYVSATFPNPLIDTLADGTPASPEBGHFEGAGIKYEPFPD 568
Qy 546 DRLNARVSPYRHKDKNAAPLNNPKKTRYALGKRWMEGVETEISGANTPKQJHAGISY 605
Db 569 GSITASV-FKLVDX-NAIVSYTAGGVVTS--GQFOVVESTFELEAKANLDEMNKALASYS 625
Qy 606 YLHSQIKTASNSRDDGIF-LIMPKSHSANLWTTYQVPE-----LTIGGVV--NAMSGITSS 658
Db 626 YTDLDITKDNANPNLIGSPVIVPAHTASLWVDYAFDTETEGSLIGGAVYQKGSWADAA 685
Qy 659 AGHAGGYATFDMAAAYRFTPLKQINADNIFNRHYARVAGANTFNIPGSERTWTANL 718
Db 686 NTLKRVSDAAVFDAAIARE-KNDWMTASVNVANVPDKEXVYKSCAGSVCGW-GDSRTITLKL 743

RESULT 27
O8UK13 PRELIMINARY; PRT; 819 AA.
AC O8UK13;
DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE TonB-dependent receptor.

```


QY 249 VLGGYLYQOHLAPYN GLPADANNKLPSPQHVFGADWNKFKONSHVFPADLKHVFG 307
 Db 413 TLOEFLYNDER---PFDGFLAFEGGIDT-PLERFFGEPDVRKVEEIGLSYREHNFN 468
 QY 308 NGGKGVKMRSDSDADSNVAFAGSKLGMKTPAGPGCGTADDAKAVGLTEIKOKLA 367
 Db 469 DNWKIRNAFRQSSDT-FDYRAEPRLNETTGILTRNFRSDDVEENYTLTDVVGK--- 524
 QY 368 FDASYSRPFRLGNTANEFIVGADVFRSTNEOGRTLLYANGGL---ALNEFRSI----- 419
 Db 525 -----FITGSIINHLLTREVLDARFISGQTOSHP-----GVTTPSINFPYNAIP 571
 QY 420 -POVDLIANARKGVKGVSHVATENLDE---FGYKSTFPADGLSLIGGRLGHYKIE 475
 Db 572 RFGDELTLN-----VVRNNQDSTGLGIFLQNIQAFADLKLVLVGRFPTVQN 620
 QY 476 SGEKTLHKASK--TKFTGYAGVYDLNDNNSLYLSQLYTPQTNLDADGLKPKRQGN 533
 Db 621 STDLRDSESSEGRYDAFPRGLIYQPIEPISLVASYSQSPQNFGRVDSIIEAERGT 680
 QY 534 QPEVGYKSYMDRLNARVSFYRMKDKNAAAP--LNPNNKTRVAAAGKRVMEGTEIIS 591
 Db 681 QYEVGYKGEFLDGLAATLAAYHTTKSNIAATDLDNDP---FLPIGEQNGIGELNVA 736
 QY 592 GAVTPKMQIHAGSYSLHSQIKTASNSRDDGIFLMPKHSANLMTTYQV---TPELTIG 647
 Db 737 GEISPGMNVIASYSHIDAEIIRNDGLQGNRPANVPFNTASFTWTELYEQDLOGGLGRL 796
 QY 648 G---VNASGITSAGMHAGYATFDMAAAYRTPPKLQIQINADINFRHRYARVAGANT 704
 Db 797 GLFYVGRDSDSSNTYI-IPGYLRTDA-ALYYKRDNRAGINIQNLFEKHYL---GANF 851
 QY 705 FNI---PGSERTWTANRYSP 722
 Db 852 GRVAIEPGLPVTYIGSFVTF 872

RESULT 29
 068590
 ID 068590 PRELIMINARY; PRT; 753 AA.
 AC 068590;
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, last annotation update)
 DE HYDROXAMATE-type ferrisiderophore receptor (iron transport protein
 DE FlU) (Probable outer membrane receptor for iron transport).
 GN PIUA OR FIU OR PA4514.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxId=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RC MEDLINE=96210657; PubMed=8633080;
 RA Ochener U.A., Vasili M.L.;
 RA "Gene repression by the ferric uptake regulator in Pseudomonas
 RT aeruginosa: cycle selection of iron-regulated genes.";
 Proc. Natl. Acad. Sci. U.S.A. 93:4409-4414 (1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=ATCC 15692 / PA01;
 RA Ochener U.A., Vasili M.L.;
 RA Submitted (MAR-1998) to the EMBL/Genbank/DBD databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Schwan W.R., Barker L., Brody L.L.;
 RA "Differences in sensitivity to PA-1806 among iron transport mutants of
 RT Pseudomonas aeruginosa compared to Escherichia coli.";
 Submitted (JUN-2000) to the EMBL/Genbank/DBD databases.
 RL [4]
 RN SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964 (2000).
 DR EMBL; AF051690; AAC06215.1; -
 DR EMBL; AF276976; AAF87743.1; -
 DR EMBL; AE004665; AAG07902.1; -
 DR InterPro: IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 KW Receptor; Complete proteome.
 SQ SEQUENCE 753 AA; 82336 MW; 4DD5430DC9D6514B CRC64;

Query Match 11.7%; Score 441.5; DB 16; Length 753;
 Best Local Similarity 22.9%; Pred. No. 7.8e-18;
 Matches 186; Conservative 129; Mismatches 318; Indels 179; Gaps 37;

QY 15 ATVLAALSSVFAQTADLETFHIKQGRSYN-----AIVTEKQDYSFVAV--TVG 63
 Db 17 ASAIGVATITAAQAQADEA---GQKTDKRVLSLDAITVGEQDETIVNDRSAS 73
 QY 64 TKIPASIREIPQSVSITITNQOVKDRNVDTPDLARKTPGLRVLSNDDGRS---VYARGY 120
 Db 74 KKYAPALDTPKVTVPVLPQVIKDTGALTADLARTTPGTFPGAGEGNPAGDRPIRGF 133
 QY 121 E-YSEYVIDGLPAQMSINGTLPLVAFPDVEYVWRGSGLPDSSGEMGGIYLVKRPPTK 179
 Db 134 NAESDTFLDG---MRVVASQTRFVFNVEQIEVSKGSGAVTGAGSTGSGISLNTLSK--TA 187
 QY 180 AFOGHAAGF--GTHKOYKAADVSGSLNSDGSVGRVMAQTVGASPRAPKNNRHETFY 237
 Db 188 KQDNFTDAGFTWQSDQYRRTLLDVRNKGIDMAAFRLNM-----KDAHV 232
 QY 238 AAMD-----WDINPDTVLG-----AGLYQOHLAPYNGLPADANKL--PSLPQ 280
 Db 233 AGRDEVSVSRWGVAPVTPFPDTPTRATLSYHLSTDDMDYGLPLTNVRSKANPSKPA 292
 QY 281 HV---FVG---ADWKKFKNMS-----HF-----VPADLKHYPGN--GGYG 312
 Db 293 SVPRDNFYGLKDRDYKSTDSGTFRIEHLNDNLTLNSSTRLVRTLLDIYVSNPDDSRG 352
 QY 313 KYGMRYSDRDADSNVAFAGSKLGMKTPAGRPGCNTADDAK-----CAVGLTEIK 362
 Db 353 NVANGYVYRKAQS-----KUSTSKGVNQITDLKANPFTGFIKHTLVTLG--- 396
 QY 363 OKALAPDASYSRPFRL-----GNTAN-EFVIGADVFRSTNEOGRTLLYANGGLALNE 415
 Db 397 --EFSYEDVHNRPYALTISGGAGNTCNARLASGDCSLNRPFGDNWTSITDGLALYT 454
 QY 416 FRSLPOVDLIANARK-----GYRGSHVATENLDEFGYKSTFPADGLSLIG 466
 Db 455 TDTKTSAAVYFTDLKLSQEWELNLGLR-----YDFD--KSSGYQTAGR-- 498
 QY 467 GRLGHYKIESEGEGTLHAKSKTKFTGY-AGAVYDLNDNNSLYLSQLYTP-----Q 517
 Db 499 GPAGYFKRE-----NNSHFMYQGLGVKPAKPNOSIYLAWSTSSNPTGETGEGQ 548
 QY 518 TNLADAGLLKPKRQNGQFEVGYKGYTMDRLNARVSFYRMKDKNAAAPLNPNNKTRRYAA 577
 Db 549 ADISVGNNGLDPERNRNLEIGTKWAFFDALSLNALFR-TDKTNARVASP-DVSTLYVL 606
 QY 578 LGRVWEGYETELSGAVTERKQIHAGSYSLHSQIKTASNSRDDGIFLMPKHSAN--LW 634
 Db 607 DGEQRVQGVELGNGLTKRWKVFQGYTYLDSIRKSYKSGEG--KNMPQTAQNNFTLM 664
 QY 635 TTYQVTPPELLIGG--VNASGITSAGMHAGYATFDMAAAYRTPPKLQIQINADNIF 691

[illegible]

QY	389	ADY	--NRFRSTNEQG	-----RT-----	TLVARGGLALNFRSPITPOVDLIAN	427											
Db	404	KFPANGFDLHNP	PGYWGNGPTDKRSARNRTELEV	KALYAFDTIALDE	-----RWDLSLG	458											
QY	428	AR-KGVRGYGH	TATENLDEFGYIGKSTH	TPADGLSLIGGRLGHV	KYKIBSGGKTLHKAS	486											
Db	459	LYDWIDGTSR	STPS	-----GKPT-----	VRADSSDG	-----485											
QY	487	KTKPTGVAGAY	VDLNDNNLSLSQLYTPQTN	-----	LDADGKLLKPRQGNQFEV	537											
Db	486	--KLSTRAGLV	FKPLENGRVYFSGT	SFNPDAEHLVTGSGVTEATGG	-LAPEKNSTYEL	542											
QY	538	GKGSYMDRLN	ARVSVFRMKDKNAAPL	PNPNKKTTRYAALGRVMEGVETE	ISGAVTPK	597											
Db	543	GTYEFLGRLE	DAALFOVKKEDAR	ELADGS	-----YVLAGEQVRGLS	SASGKLNH	598										
QY	598	WQIHAGYSYLH	SQIKTASNRDDGIFL	-LMPKHSANLWTTV	YTPTELTIGGVNAMS	--G	654										
Db	599	WDLFGTYTYL	DSSETLKS	PNPREGALGNT	PPRSLSLWSTYEL	DEGVTLGARYGVYSORN	658										
QY	655	ITSSAGMHAGG	YATDANAAREFP	PKLQINADINFRHYARV	699												
Db	659	VTSVDNGKLD	AYVWHNMMLGYQ	TRDLKLQNLNLD	FKAVVERV	703											
RESULT 31																	
ID	QY1422	PRELIMINARY;		PRT;	732 AA.												
AC	QY1422;																
DT	01-MAR-2001	(T=EMBLrel. 16, Created)															
DT	01-MAR-2001	(T=EMBLrel. 16, Last sequence update)															
DT	01-OCT-2001	(T=EMBLrel. 18, Last annotation update)															
DE	Probable TonB-dependent receptor.																
GN	PA1322.																
OS	Pseudomonas aeruginosa.																
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;																
OC	Pseudomonas.																
OX	NCBI_TaxID=287;																
[1]																	
RP	SEQUENCE FROM N.A.																
RC	STRAIN=ATCC 15692 / PA01;																
RX	MEDLINE=20437337; PubMed=10984043;																
RA	Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,																
RA	Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,																
RA	Garber R.L., Goltry L., Tolentino K.R., Kas A., Larbig K., Lim R.M.,																
RA	Brody L.L., Coulter S.N., Folger K.K.-S., Wu Z., Paulsen I.T.,																
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Olson M.V.,																
RA	Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,																
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an																
RT	opportunistic pathogen."																
RL	Nature 406:959-964(2000).																
DR	EMBL; AE004561; AAG04711.1; --																
DR	InterPro; IPR000531; TonB_boxC.																
DR	Pfam; PF00593; TonB_boxC; 1.																
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.																
KW	Receptor; Complete proteome.																
SK	SEQUENCE 732 AA; 80835 MW; C2171C0PDC409FBB CRC64;																
Query Match																	
Best Local Similarity 23.3%; Pred.No. 2.5e-17;																	
Matches 178; Conservative 123; Mismatches 299; Indels 165; Gaps																	
QY	17	VLAALSSVFAAQTA	-----DLETVHKGRS	YNAIVTEKNGDYSSFAVTVGT	KIPASL	70											
Db	22	LMACSSPALAEAR	KDDPLELGADTVTGEQASSR	VERSASAKYA	-----VP--L	70											
QY	71	REIPQSVSIITN	QOQVDRNVDT	FDQLAKT	PGLRVLSNDG	---RSSVYARGVE-YSEYN	126										


```

Db 131 VDGLRDSAQOT---SRSDLFNLEAVEVITKGPNSVFEGAGTSGSINMVSKOPKAEPHTLIG 187
Qy 187 AGFGTHKOYKAEDVSGSLNSDQ---SVRGVMAQTGASPRPAEKNNRHETFYAAADMD 243
Db 188 AGLGTRVYRLTLDTNQLLEGESTAFRLNMAHNDVAGEQGLDKOR-----WG 238
Qy 244 INPDVVLG-----AGYLVOQRHLAPYNGLPADANKLPSLPQHVFVADWKNF---K 292
Db 239 IAPSLTFGLDPTPTLLTSFYHQRDNDLDPYGVFALNGRKLGVSHRDYFG--WRMLDEER 296
Qy 293 MNSHDFADLKHVFGNGYGVKVMRYSDDRADQ---SNVAFASGKLGMKTPAGRGNT 347
Db 297 IDNVATLKLHEBDDQLOQLNRLYSHLRDTVYSASHNOKGPPGGRYLPAG--PQAYG 355
Qy 348 ADDKA-----CAVGL-----GTEIKOKALAFDASYSRPRPLGTANEFYIG 388
Db 356 RDSKTRMWINQTNLGRFDTGLAHTLLIGFELSRE-----TYR-----TTSYMLG 403
Qy 389 ADY--NRPRSTNEQG-----RT-----TYARGGLALNEPRSTPQVDLIAN 427
Db 404 KFYANGFDLHNPFGYNGPPTDKRDSARNRTELEVKALYAFDTIALDE----RWDLSLG 458
Qy 428 AR-KGVREYSHVATENLDEFYIGKSTFHPADGLSLIGGRLGHYKLESGBGKTLHKAS 486
Db 459 LRYMDISTSKSTBS-----GKPT-----VRADSSDG----- 485
Qy 487 KTFETGVAGAVYDLNDNNSLYLSQLYTPQTN-----LDADGKLKPRQNOFEV 537
Db 486 --KJSTRAGLVFKPLENRVYFYSYGSFNPAEHLVTTGSGVTEATGG-LABEKSTYEL 542
Qy 538 GYKSGYMDRLNARVSFYRMKDKNAAAPLNPNNKTRRYAALGKRVMEGETEISGAVTPK 597
Db 543 GYKMEFLGRRLLEPAALFOVKKEDARERLADGS---YVLAGEQVRGLERLSASGLMEH 598
Qy 598 WQIHAGSYLHSQIKTASNSRDDGIFL-LMPKHSANLMTTYQVTELTIGGVNAMS--G 654
Db 599 WDLPAITYILDESETLKSSNPREGALNTPRSLSTLWSTYELPBGVTLGYGARVVSQRN 658
Qy 655 ITSSAGMHAGYATFDAMAAYRFTPKLKIQINADNIFNRHYARV 699
Db 659 VTSVONCKLDAYVWVNMMLGYQATRDLTQLNLNDLFPKAYVERV 703

RESULT 32
Q9FOF9 PRELIMINARY; PRT; 755 AA.
AC O9FOF9;
DT 01-MAR-2001 (Tremblurel. 16, Created)
DT 01-MAR-2001 (Tremblurel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblurel. 19, Last annotation update)
DE Putative ferrichrome iron receptor PhuA.
GN PHUA.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21064758; PubMed=11136137;
RA Galindo M.A., Day W.A., Raphael B.H., Joens L.A.;
RT "Cloning and Characterization of a Campylobacter jejuni Iron-Uptake
Operator.";
RL Curr. Microbiol. 42:139-143(2001).
DR EMBL; AF241226; AAC47637.1; -
DR HSSP; P06971; 1BY5
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 755 AA; 83851 MW; C4140C3D258202E0 CRC64;

```

Query Match 11.4%; Score 432; DB 2; Length 755;
 Best Local Similarity 23.9%; Pred. No. 2.8e-11;
 Matches 179; Conservative 116; Mismatches 331; Indels 124; Gaps 31;

```

Qy 15 ATVLAALSSVFAQTADLETVNHKGGRSYNAIVTEKNGDYSSFA--VTGKTKIPASLRE 72
Db 18 AVVLAATCHGVQAD---DQVTLPSQVTGELDRPQGPDPGYAKESLTLTSTKSTPLSE 74
Qy 73 IPQSVITITNOQVDRNVDTFDQARKTPGLRV--LSNDGRSS--VYARGYSEYNI- 127
Db 75 TPRSVSVTRKRIBEDQSKSLTLDVLYGVPGIFAPFPFAAGDGLAGDLFFIRGFNATDYGVG 134
Qy 128 ---DGLPAQOMSINGTLNLFAPDRVEVMRGPGLFDSGSGMGIVNI-VRRRPTAPQGH 184
Db 135 LLRDLRVQGNRYD-TISEPYGLERVEFRGPTSLIGENAPGGLVNLVSRPPTAATQGE 193
Qy 185 AAAGFTHKOYKAEDVSGSLNSDQSVYRGVMAQTGASPRPAEKNNRHETFYAAADMDI 244
Db 194 VQLGSGNDRQQLNLDVSGPLNDGNNVIGRV--MVGSEATQVDHVDQDDIYLAISLTL 251
Qy 245 NPD-----TVLAGYLYVOQRHLAPYNGLPADA---NNKLPSPQHVFVG-ADWKNFKRNS 295
Db 252 NFDVNSLTLLST---YQKDRTKLELGLPAGTLLRNPNKIDKDTFLGNKDMWTFEREV 308
Qy 296 HDVFAIDLKHVFGNGYGVKVMRYSDDRADSNVAFASGKLGMKTPAGRGCTADDAKAV 355
Db 309 WTLGYEFTFRNDWMQFRNSRYMQSRIQRNEIWPGA-----LNNA----- 349
Qy 356 GLGTEIKOKA-----LAFDASYSRPRPLGTANEFYGADYNRPRSTNEQRTTLY 406
Db 350 GFGTNVSTQAYDRNKSTITYSLDNQBEGKFTGLAQHTLVAGASFDR-TSRNQD-----W 403
Qy 407 ARG-GLALNEFRSI---PQVDLIANARKGVGYSHVATENLDE--FGIYGKSTFHPAD 459
Db 404 SAGGSPSINVENPWTSEPTTPF-----TIONSQLOQWAGLYSQLSQSKYEN 450
Qy 460 GLSLIGGRRLGHYKLESGBGKTLHKASKTK-----FTGYAAVYDLNDNN 504
Db 451 WIFLLGQ-----RFDVDSQYRNKFDATSPVSNASDLDYTDDBFTWQAGVMYQFENG 504
Qy 505 SLVLSLSQLYTP-OTNLADADGKLKPRQNOFEV--YKSGYMDRLNARVSFYRMKDKN 561
Db 505 SPVYSYSFPPVQQTTSASGP-LDPTTAEYEGVGLKTEPRGKNTMPTAAV-FDLRKEND 562
Qy 562 AAAPLNPNNKTRRYAALGKRVMEGETEISGAVTPKQIHAGSYLHSQI-KTASNSRSD 620
Db 563 VYFPAQIND---YRQVDSRSKGVLEINSDISANLVTAAITYTDARITKDAIPGSLLE 618
Qy 621 GIFLL-MPKHSANLMTTYQ---VTPELTIGGVN-----AMSGITSSAGMHAGYATF 669
Db 619 GNMQTVGPRNQASIMANVRFDLGALRGIRLGQGVRRHPTFAVTAESLYGKLDGDTLV 678
Qy 670 DAMAAYRFTPKLKIQINADNIFNRHYARV 699
Db 679 DALIGYDIDENWSVDLNKKNFFDQGVGRGL 708

RESULT 33
O86424 PRELIMINARY; PRT; 732 AA.
AC O86424;
DT 01-NOV-1998 (Tremblurel. 08, Created)
DT 01-NOV-1998 (Tremblurel. 08, Last sequence update)
DT 01-DEC-2001 (Tremblurel. 19, Last annotation update)
DE Ferrichrome iron receptor.
GN PHUA.
OS Pantoea agglutimans.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K4;
RX MEDLINE=98348446; PubMed=9683481;
RA Killmann H., Herrmann C., Wolff H., Braun V.;
RT "Identification of a new site for ferrichrome transport by comparison

```


QY 669 FDMAAARFTRKLCQINADNIFNRHYYARVGANTENIPGSERTWTANLRYSE 722
DB 809 TDAISIFYK-RDNMRALNFKNLFDTKYEE---SOSFFIVPAPFVLGTVEFEF 858

RESULT 35

Q926C6 PRELIMINARY; PRT; 724 AA.
ID Q926C6
AC Q926C6
DT 01-DEC-2001 (TREMblrel, 19, Created)
DT 01-DEC-2001 (TREMblrel, 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel, 20, Last annotation update)
DE Putative ferrichrome-iron receptor precursor protein.
GN FH042 OR R02269 OR SMC01657.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
ON NCBI_TaxId=382;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstead P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Golliau A., Kahn D., Kiss E., Lelaure V., Maury D.,
RA Pohl T., Portetle D., Puhler A., Purnelle B., Rampeger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.,
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021".
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC46848.1; -.
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 724 AA; 79240 MW; CB919FBBE9794A9C CRC64;

Query Match 11.2%; Score 423; DB 16; Length 724;
Best Local Similarity 23.5%; Pred. No. 9e-17;
Matches 189; Conservative 112; Mismatches 361; Indels 122; Gaps 27;

QY 5 MSVRIMNTAATVLAISVFAAQTDLFTVHIKGRSYNAIVTEKNG-----DYSSF 58
DB 9 LPLRLRLAGTSALALVATA--QAQAEQETVSGNDSTALETLVNGSGGVITAEAGVGT 66
QY 59 AVTVGTIPASIREIPQSVSIITNQVDRVVDFTDQARTPGARV--LSNDGRSSVY 116
DB 67 SSATGAKIDTPFLETPOSISTVTQOLKDNRPQTLETLATTPGTRVGAYGDFDPDAFF 126
QY 117 ARGYE--YSEYNIDGL--PAQMOISINGTLPLNFAFDREYVVRGPGSLFDSGEMGIVNL 172
DB 127 VRGFDVTVYSGVFRNLQPAVDSIFKNEP--YGLEGVSIIRGSSALYKATGAGGLYNL 184
QY 173 VRKPTAFQGHAAAGFTGHKOYFAEADVSGLSNDSGVRGVN-----AQTVGASPR 225
DB 185 ITRKPTEDTLREYVOVYQSHDRYQGFEDFSGPVNENDPVYRLTGLRLADDTQGVGLADD 244
QY 226 PAKNNNHETFFYAADMDINP-----TVLGAGLYLXOORHLAPFNGLRPADANNKLPSLPQ 280
DB 245 RA-----YIAPATFWKDEDETKLTVLGE--YSRTN---SGGATATYVNDPLTGEAT 289
QY 281 HVEVVG-ADWVKFKNNSHHVFADLKHVFGNGGYGKVMRYSDRADSNYAFAGSKLGMKTP 339
DB 290 DIFAGNDFPDSVQKQKRGVGEFEHRLNDTFVFRONARVSTLINDADMAFYAP----- 343
QY 340 AGRPGCTADDKACAVGLGT-ETQKALAFDASVSRPRLGNTANEVIGADYNRFSTN 398
DB 344 -----NAAPPTLLDSSAGTYDERLTAFFVIDNQLAEKFDGALHTHTLLAGDVYTKLFR 397
QY 399 EOGRTTIYANGGLANFRSIPQVDLIANAKRGVSHVATENLDEFGIYGSTHRA 458
DB 398 LDGKGVSPPLDTKTPQGRPVDAIDP-----NRTVADQWQGLTYLQDQIR-Y 444

QY 459 DGLSLIGGRGLGHYKISGECKTLHAKSKT-----KFTGYAGAVYDLNNSLYLS 509
DB 445 DAWTLTVGR---YDWMSTDTDTMDLTDLSLTVYSQDKERSGRIGLTYQDFGLAPYIS 501
QY 510 LSOYTPQTNLDAD-GKLLKPRQGNQFEVYKGSYMDRLNARVSFRMKDKNAAPLNP 568
DB 502 YSTAFAPNAGINKETNOPEKPTBEOOEIGVKYLLPNSNTLITLALFNIDQKGLY-LEA 560
QY 569 NNNKTRVYALGKRVMEGVETISGAVTPKMOIHAGYSYLSQITASNRDGIFFLMPK 628
DB 561 SGDTAIVQYQKRLASRGEIENANSLDNGISLASYATVYKTIQGSRTIGNYSSAPQ 620
QY 629 HSANLWMTTYQVTPP-----LTIGGVNAMSGITSSAGMAGGY-----ATFDMAAYRF 677
DB 621 HMASIMAHYTL-PEDGPFYFSLGGARF---VGSYGNPDNTKNSRVLFDSVGVDF 676
QY 678 T-----PKLQINADNIFNRH-----YARVGANTFNIGSERTWTANLRYSE 722
DB 677 AAIDQKYEGLHLQVNAATNLPDRREAVCTAGYCHRDQ-----RTVIGSIRYMW 724

RESULT 36

Q82908 PRELIMINARY; PRT; 696 AA.
ID Q82908
AC Q82908
DT 01-MAR-2002 (TREMblrel, 20, Created)
DT 01-MAR-2002 (TREMblrel, 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)
DE Periloxamine B receptor.
GN STY0396.
OS Salmomella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmomella.
ON NCBI_TaxId=601;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jorgels K.,
RA Kosch A., Larsen T.S., Leather S., Moutie S., O'Garra P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RT "Complete genome sequence of a multiple drug resistant Salmomella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627266; CAP08819.1; -.
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC.1.
KM Receptor; Complete proteome.
SQ SEQUENCE 696 AA; 77130 MW; E3FCEA55PD723FF CRC64;

Query Match 11.2%; Score 421.5; DB 16; Length 696;
Best Local Similarity 23.1%; Pred. No. 1e-16;
Matches 175; Conservative 121; Mismatches 353; Indels 109; Gaps 25;

QY 12 MTATVLAAL-----SSVFAAQTADLETYHIKGRSYNAIVTEKNDYSSFAVTVG 63
DB 1 MPATTRALLIGALIGATFPFLAQETTKNDV-----IVT-----SPVSGA 43
QY 64 TKIPASIREIPQSVSIITNQVDRVVDFTDQARTPGRLVLSNDDGRSS---VYARG 119
DB 44 TRLATPDIETPOSVSITTRQOFEEQATSVQAVSYTPG--VYSNQIGASNRDPDYIYLRG 101
QY 120 YSESEVY---IDGLP--AQMOISINGTLPLNFAFDREYVVRGPGSLFDSGEMGIVNLVR 174
DB 102 FSDGSLDNVYLDGKMMGDNSSHSLVVDWPFLEDIHVAGPASYLYGRSSPGGIYSLTS 161
QY 175 KRPTKAFQGHAAAGFTGHKOYKAADVSGLSNDSGVRGVNMQVAGASPRPAKNNRHE 234


```
Db 162 RKPAPDAGGEVKLFAGNNQRGAAPDVTGPTDNNRVAARLSGWTYADSQFTPLKEERY 221
QY 235 TFYAAADWDINPDVTLG-AGYLYQ-----ORHLAPYNGLPADANNKLSLPQH 281
Db 222 ALMPSLTWRTITDRTRLDLMAYLHRDPGGSHSGLPYQGTWVPYNG-----GKISNT 272
QY 282 VVFVGA--DWNKFKMNSHDVADLKHIFGNGGYGKVGWRYSDRDADSNYAFAGSKLGMKTPA 340
Db 273 FPEGEDDYDKYDRRENKMGVNIHLEFDNGMSVRQKRLYLHTKVTLMQVYAAAGLN----- 327
QY 341 GRPGCMTADDKACAVGLGTEIKKALAFDASYSRPRFNGTANEFVIGADYNRFRSTNEQ 400
Db 328 -----ETALNRGYS-----GSGEKMSAIALDNQLDGSVDTGAINHRLVAGIDYQ-----DRS 374
QY 401 GRTTLYARGGLALNEFRSIPQVDLIANARKVGRVGYSHTVATE--NLDFEIGYKSTFHFA 458
Db 375 NHTTGY-----YGAFPPIDAFNPVYGAQPDYITLRSREKHLRQGTGYLQDQM-SW 424
QY 459 DGLSLTGGRLGHYKESGECKTLHKA-----SKTKGTGAGAVYDLNNDNSLYLSLSQLY 514
Db 425 DRWRFTLGGRYDRSVSNID--KLHDSRSLDKNNVSTRAALLYLFDNGVAPYLSYSTAF 482
QY 515 TPOTNLADGKLLKPRQGNQFEGYVYKSGYMDRLNARVSFYRMKDKNAAPLNNKTR 574
Db 483 TSTSFADENGVNLEPMKKGWEAGVKEPLGNSQFSAAVYRINQINATKEEPTDP--- 539
QY 575 YAAALGRKRMVEGTEISGAVTPKQIHAGYSYLHSHQIKTASNRDDGIFILMPKHSANLW 634
Db 540 YRSIGETESKGELEAISHLSDSVRLQAAVYTDIIRYKSSPQEQGKRAVYAPRNOASAW 599
QY 635 TTQVTPTE-----LTIGGVNAMSGITSS-AGMHA-GGYATFDMAAAYRTP-----KLKQ 684
Db 600 LSYDVKSGLELWLTGSGIRYVNGVTSRDLNTHLTPSYTTLVDMVVGVDLSIGLNGLSAQ 659
QY 685 INADNIENRHYARVGGANTFNIPGSRBTWTANLRYSF 722
Db 660 LNVNLTDKRYVACNSL-SYCYFGEAERSIVGSWSAF 696

RESULT 37
Q8YU26 PRELIMINARY; PRT; 853 AA.
AC Q8YU26;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN AUR2185.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RX SEQUENCE FROM N.A.
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120."
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73884.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete Proteome.
SQ SEQUENCE 853 AA; 94548 MW; 67EA77C09E049437 CRC64;

Query Match 11.1%; Score 420.5; DB 16; Length 853;
Best Local Similarity 23.1%; Pred. No. 1.6e-16;
Matches 187; Conservative 117; Mismatches 327; Indels 177; Gaps 30;

QY 3 QFMSVFRINMTAATVL--AALSSS-----VF--AAQTADL-----ETVHIKQRSYN 45
```

```
Db 135 QDANTIRSVIGETILTAELEFDSDEGLIFGLTSVTASLPQPQETQSEQSPAAEEPIE 194
QY 46 AIVT-EKNGDYSSFAVTVGTIKIPASLREIPOSVSIITNQVKDRNVDTFDOLARKTPGLR 104
Db 195 LVVTGEKDGQVQVNTATV-TRTNTPIIDIPOSIQVPPQVLEDDQOITRVDDALRVNPGVL 253
QY 105 VLSNDGSRSSVVARGYEYSEYNI--DGLPAQMOSINGTLPNLFAFDRFVEMVRGSGLFDS 162
Db 254 GSTNAFIGNQITIRGFTSNLPILLRDPRIYE--NFSQETSNLERIEVLKGPASVQYG 310
QY 163 SGMGGIVNLVRKPTKAFQGHAAAGFTHQYKAEADVSGSLNSDGSVGRVMAQTGA 222
Db 311 QLDPCGVNLVTKKPLSEPFVEIOAQFGSGLIRSFDSVSGPLTDDGKLLYRLNA-TYQR 369
QY 223 SPRPAEKNNRHETFAAAD--WDINPDVTL--GAGLYLQQRHLAPYNGLPADANNKLPSL 278
Db 370 BEGPRDFNTETETEFPIAPSLTWKISDRTNVDFSLBYLSTR---PFDTSLVAFGRSVADV 426
QY 279 P-QHVVFVGDWNNKPMKNGHDVPAFLKHVFGNGGYGKVGWRYSDRDADSNYAFAGS- 332
Db 427 PYSRVNDPD-DPIDTKSFSIAYNLEHREFSDNWTLRNSFRYLQQDLFTQATLAGSLNETT 485
QY 333 KLGMKTPAGRPCGNTADDKACAVGLGTEIKKALAFDASYSRPRFNGTANEFVIGADYN 392
Db 486 GILTRTYAQR-----EYKSDDYSLQTNVVGKFTTGSIKKHTLLAGVDEN 528
QY 393 R-----FRST-----NEQGRTTLYARGGLALNEF 416
Db 529 RGLLDDLVRGRTTLNIPNVPVGPRTDFSTLPPATPFKNETTLRGFLYLOQQTALNN- 587
QY 417 RSIPOVDLIANARKVGRVGYSHTVATENLDFEIGYKSTFHPADGSLIGGRLGHYKIES 476
Db 588 ----QFTVLGRLYDVTDFPKDTDESKVD-----SAWSPVIGL----- 622
QY 477 GEGKTLHKASKTKFTGYAGAVDYDLNNDNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQFE 536
Db 623 -----VYKPVENLSIYTSYRSFVPSFSDANGDFLOPERGAGYE 662
QY 537 VGYKGYMDRLNARVSFYRMKDKNAAPLNNKTRVAALGKRVMEGVETEISGAVTP 596
Db 663 VGKAELELQNLFATLAYFDITKQNVST-ADPDVLGAS-VATGEQKRSGLIELSAGVTIAP 720
QY 597 KWOIHAGYSYLHSHQIKTASNRDDGIFL-----LMPKHSANLWTTTQV---TPELTIGG 647
Db 721 GWNIIAGYAYTDAEI-----SKDNTIAGVNRLEPGTPKHSANLWTTVEIQKGSILQGLGFL 775
QY 648 GVNAMSGITSSAGMHAGGYATFDMAAAYRFT-----PKLKLQINADNIENRHYARV 699
Db 776 GVNIV-----GKRFNGFQNDPEVDSYFLTNAALFYRQNNRVRVGLNFNIFDINY---I 825
QY 700 GGANTFN-----IPGSERTWTANLRYSF 722
Db 826 SSAATLTRTSIEPQGPFTVVGSIISLEF 853

RESULT 38
ID P72609 PRELIMINARY; PRT; 853 AA.
AC P72609;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FHUA OR SLR1490.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RX SEQUENCE FROM N.A.
RA MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiura M., Sasamoto S., Kimura T.,
```


RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nartuo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.,
 RT "Sequence analysis of the genome of the uncultured cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-113(1996).
 DR EMBL; D90899; BAB1609.1; -.
 DR HSSP; P06971; 1BY5.
 DR InterPro; IPR001484; Pyrokinin.
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB boxC_1.
 DR PROSITE; PS00539; PYROKININ; UNKNOWN_1.
 KM Complete proteome.
 SQ SEQUENCE 853 AA; 94170 MW; 11369D592F1F9D9C CRC64;

Query Match 11.0%; Score 416.5; DB 16; Length 853;
 Best Local Similarity 23.2%; Pred. No. 2.7e-16;
 Matches 178; Conservative 116; Mismatches 311; Indels 161; Gaps 27;

QY 10 INMTATVLAALSSVPAQAQDADLETVHIKGRSNAIYT-EKNGDYSSFAVTGKIPA 68
 DB 165 VNATDQSLVIGLSPGKGVADDEB-----GNDAIQVVTGEDEGAVDADATATLTDT 217
 QY 69 SUREIPQSVIITNQVKRNVDTFDQARKTPGLRVLSNDGRSSVYARGVESEYNID 128
 DB 218 PLRDIPOSIQVPOQVLEDRDIIIRASEALQNVSGVQRGVTGQTSSEI-----FNIR 268
 QY 129 GLPAQWQSIINGTLPLNLFAR-----DRVEYMGPSGLFDSGEMGIVNLVKKR 176
 DB 269 G-----FQGGGGLRDFKRDNFSPIDTNLQIETVLKGPASVLYGNLDPGVINITYIQ 324
 QY 177 PTKAQGHAAAGRGTHQYKAADVSGLSNDSGVARGV-MAQTVGASPPRAKNNRHET 235
 DB 325 PLSEPFYEVAQAGNGLVRPTIDLSGLNSQRTALYRLAAVEGSGNRDDEPTEVARPF 384
 QY 236 FFAADWDINPTVYGA--GYLYQOQHLPYN-GLPADANNTPLPQHVFVGADWKKER 292
 DB 385 ISPVTWQISDQTDLRFEWDLYLDR--PFDRGIYA----- 418
 QY 293 NMSHVDFADLKHYPNG-----GYKGYKMYSPRDADSNYAFAGSKLGMKTTPAGRPGCNTA 348
 DB 419 -----FGTADIPDFRVLGLDLPDARTNFS-AGYRLERH-----FS 455
 QY 349 DDKACAVGLGTEIKOKALAFDASYSRPFRL-----GNTANEF-----VIGA 389
 DB 456 DNWK-----LRNRFPSYLDQAAEQTELVLDDETTGNLSKQFSRNEQOINNELQTDLIG- 510
 QY 390 DYNRFSTNEQRTLYA-----RGLALNEFRSIPVDLI-----ANARKGV 432
 DB 511 ---KF-STGPIQHTLLFGVDLSWQSAPIFRGVA-----APTINIFPVYGVYARPSI 560
 QY 433 RGYSTTATE-NLDFEGYIGKSTFHPADGLSLIGGRLGHYIESGEGTLHKASKTKEF 491
 DB 561 NDFPVSFSEGGTNTLIGFLDQVTLTDLTKLIMGGRPTIIDQSSSSNDSSEDERYDQAFS 620
 QY 492 GYAGAVYDLNDNNSLYLSLSQLYTPQTNLDADGKLLKPGQNGQFPGYGVKSGWMDRLNAR 551
 DB 621 PRLGIVYQPIEIVLSYASFSRSFQNFGRFDSGLDEPFGQYEVGANGFLLDGRLLAN 680
 QY 552 VSFYMKDKNA--PLNPNKKTRYAALGKRMVEGVETELSGAATPKQKHAGSYLHS 609
 DB 681 LAAVEITVSNLAVTDPENPNFS-----IPSGEQRSGVEFDINGELLPGNNIIASVAYIDA 736
 QY 610 QIKTASNRDQCIPLMKRHSANLWTTQVTPETLITGGVNAMSGITSSAGNHAGYATF 669
 DB 737 RVTYKDNLEPGLNLEGVFNSSLSMSTYEQ-----AGDLQIGLGLGLFYVGERQG 788
 QY 670 DAMAAYRFTPKLKTQINDNIFNR--HYARVAGANTFNIPESERT 713
 DB 789 DLNNSFOIPLYRTDI---SVFYRRNNRRAINNNLNFIDYTEAT 831

RESULT 39
 ID 08YTV3 PRELIMINARY; PRT; 857 AA.
 AC 08YTV3;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Ferichrome-iron receptor.
 GN AL2610.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 NC NCB1_TaxID=103690;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.,
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 DR EMBL; AP003590; BAB74309.1; -.
 DR InterPro; IPR000531; TonB boxC.
 DR Pfam; PF00593; TonB boxC_1.
 KM Receptor; Complete proteome.
 SQ SEQUENCE 857 AA; 94906 MW; 7D4C98620AEB8007 CRC64;

Query Match 11.0%; Score 415.5; DB 16; Length 857;
 Best Local Similarity 23.6%; Pred. No. 3.2e-16;
 Matches 174; Conservative 106; Mismatches 296; Indels 161; Gaps 28;

QY 47 IYTEKNQDSSFAVTVGKIPASLEIPQSVIITNQVKRNVDTFDQARKTPGLRVL 106
 DB 204 VVTGQDQDRYSVNAATATRTDTPRLDIPQSIQAVPROVLEDRQVIRASDARSVSGVQGG 263
 QY 107 SNDPDRSSVYANGYSEVNIIDGLPAQWQSIINGTLPLNLF-----AFDVEVNR 154
 DB 264 NRYGQTSSEV-----FNIRGFP-----QFGNLRDGFNNNNSIVETANLERIEVLK 310
 QY 155 GPSGLFDSGSEMGWGIIVLKRKPTAFQGHAAAGF--GTHQYKAADVSGLSNDSGVSR 212
 DB 311 GPASVLYGNLDPGVYINVTYQPLS--EPFYAAGIQVSGFGLVRYTLDSGLNPERTL 368
 QY 213 GRVMAQTVGASPPRAKNN--RHETFYAA--ADWDINPTVYAGYLYQOQHLPAPYN-G 266
 DB 369 YRLNA-----AYERGNFNFNPFTEVERFISPVYTWKIGDRTDLRLLEY-SNDKRPYDRG 423
 QY 267 LPADANNTPLSPQHVFVGADWKKRKNMSHVPAD-----LKHYPNGGYKGVKGR 317
 DB 424 LVAFGTG-IADIPFDRVLG-----EPDFFSRITFLAGYRLERHFNDFDKLRNFR 473
 QY 318 YSDRPADSNYAFAGSKLGMKTTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 377
 DB 474 YSSSDQTN-----RLPGRFLNETTGE-----LRREBS 501
 QY 378 LGN-TANEFVIGADYNRRSSTNEQRTLYA-----RGLALNEFRSIPVDLIANAR 429
 DB 502 LGEEFTVRYVELQTDLVGNFATGSIQHTLLFGVDLSWVNDGVSLL-FEPAPASIN-IFNPV 558
 QY 430 KGVRGYSTTATENIDER-----GIYKSTFHPADGLSLIGGRLGHYKLES 476
 DB 559 YGI-----ATRPSRDEADVPPFGSQTDSIGVFVQNTILAEMLIKLLVGRFNDIDQSS 612
 QY 477 GEGKTLHKASKTKFTGYGAVYDLNDNNSLYLSLSQLYTPQTNLDADGKLLKPGQNGF 536
 DB 613 ASDERQDA-----FSPRGIYVQPIEIVLSYTSFSRSFQNFGRFADSSLEPARGIYE 668
 QY 537 VGYKGYMDRLNARVSFYRMKDKNA--PLNPNKKTRYAALGKRMVEGVETELISGAV 594
 DB 669 VGVGRGEFLNGSLITNLAAVEITRSNLAVTDPDNPNFS-----IPSGEQRSGRVEIDVTGQI 724

Wed Dec 18 08:35:28 2002

```

595 TPQWQHAGSYLHSGIKTASNSRDDGIFLLMPKHSANLWTTQYVTPETLTGGGVNMSG 654
725 LFGWNLIAATYATDARVTKDDNLQPCNLLDGVFNSASLWYEQ-----TCDFQG 776
655 ITSSAGMHAGG-----YATDAAARVFTPKLKLQKLNADNINRHHYAA--- 697
777 LGFGLGLFVYGERQGLNNSFOLPSYVRTDASIFYR-RNNWRAGINNNLNFVNDYIEDSG 835
698 -----RVGGANTFNIPGS 710
836 QRRNRINPGEPTVRGT 852

RESULT 40
Q8XVU8 PRELIMINARY; PRT; 689 AA.
AC Q8XVU8;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative ferrichrome-iron receptor protein.
GN RSC2729 OR RS00125.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OC NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RC MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunin S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RT Nature 415:497-502(2002).;
RL EMBL; AL646071; CAD16436.1;
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 689 AA; 74213 MW; 6E9B8AFCD092C5 CRC64;

Query Match 10.9%; Score 412; DB 16; Length 689;
Best Local Similarity 22.8%; Pred. No. 3.7e-16;
Matches 170; Conservative 127; Mismatches 330; Indels 120; Gaps 26;

QY 20 ALSSVFAQAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQSVI 79
19 ALBEAVVRAQAGE-----RLRARATSATLTD-----TPLKDVQPSGV 57
80 ITNQVQKDRNVDTFDQARKTPGLRVLSNDDGRSSVVA-RGY-----EYSEYNIDGLPAQ 133
58 VTRAALDGFATRLDALTALDWMVSGISRONLGGIADNFAGFAGDNLNLTGSDYLVNFGSA- 116
134 MQSINGTLPLNLPFADRVVVRGSGSLGSDSGEMGGIVNLVVRKPTKAFQGHAAAGFTHK 193
117 -NRANSVPVDVTINARIIDLVLKGSAAALYGRSDPGGIVNITRTPQKPSREITLAGSHD 175
194 QYKAEADVSGSLNDSGVRGVRVMAQTGASPRPAEKNNRHETFYAADWDINP----- 246
176 QYRLATELTGSLGSH-----FAYRLGLA-----AENNHGFPFSASRRYIAPFTWLPT 225
247 -DTVLGAGLYQQRHL-APYNGLPADANNKLPSPQHVFGADWN-----KPKMNSHDYFAD 301
226 DDTVV--TYAEAAQLRAPFDRGIVAINRQLGALPDSFLGEPGGDTTTRVTSQSHL--S 281
302 LKHVFGNGYGVKVMRYR-----DEDADSNVAFAGSKLGMKTPAGRPCNTADDKACAVG 356
282 VEHLQSGMKVDAGLSYRSLSFGKSSDASRLDDGRTLWQARED--YHANDLAGRVD 339
357 LGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRRFRSTNEQGRITLYARGGLALNEF 416

340 LQGEV-----ATGVRHTLVAGADFYNFRDVPVLYRANPSASAPYADIF 384
417 RSI-----PQVDLIANARKGVGYSHYVATENLDEFGYKSTPHPADGSLIGGRLGH 471
385 DPVYQPPALRPPSSITRESORG-----FGAFVQDQVTLTPQWKLLAGVRMDR 432
472 YKIESGEKTLHKAS--KTKFTGYAGAVYDNDNNLSLSLSQLYTPOTNLDADGKLKP 529
433 FLQHADNLTGARVAQOQTAVSPRLGLVYQPTQTLISLYANTARSFRPNTGVAQGNAPAP 492
530 RQGNFEVGYKGSYMDRLNARVSYRMMKDNAAA--PLNPNKKTRYAALGKRVMEVGE 587
493 ERGRGEVGALETADGRFGGTLDALYSIDKTNVLTGDSRDPVQFRTAGAVRSR-----GVE 548
588 TEISGAVTPKQIHHAGSYLHSGIKTASNSRDDGIFLL-MPKHSANLWTTQYV---TPE 642
549 LDVSGQLTPNLKVLGTYYATDARV-TADTVLPSGAPLSNIPHSASALGLYEFAGSLGR 607
643 LTIGGG---VNAMSGITSSAGMHAGGYATFDMAAYRFTPKLKLQKLNADNINRHHYARV 699
608 AGVGGGVVYGERAGNSTDNGFKLPAYATVRLNGVYQPTRALRLSLTIDNLFDKRY--- 664
700 GGANTFN-----IPGSERTWTANLRYSF 722
665 --ASSYNELMWAFGAERQVTLAATYTF 689

RESULT 41
Q92YCI PRELIMINARY; PRT; 714 AA.
AC Q92YCI;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Putative ferrichrome-iron receptor.
GN RA0963 OR SWA1747.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OC NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RC MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RA "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
DR EMBL; AE007282; AAK65621.1;
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 714 AA; 78235 MW; 4DAB5F4862934A69 CRC64;

Query Match 10.9%; Score 411; DB 16; Length 714;
Best Local Similarity 24.6%; Pred. No. 4.5e-16;
Matches 187; Conservative 110; Mismatches 344; Indels 120; Gaps 30;

QY 12 MTAATVLAALSSVFAQA-----TADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
24 LLGCTAFFALTALPSAQADAVPEGDTTIVTIVAHGAGGSGVLNLTDEDSKIATETGA 83
65 -KIPASLREIPQSVSIITNQVQKDRNVDTFDQARKTPGL--RVLSNDDGRSSVYARGYE 121
84 GKMPDILIPASVSVITSKEIBEAAADTIEQVQVVTAGVVTDFYGSDDRDYDFDINGFT 143
122 YSEYNIDGLPAQMSQINGTLPLNLPFADRVVVRGSGSLGSDSGEMGGIVNLVVRKPTKAF 181
```



```

Db      144 PYRR-DGL-AIGRTFGRVEEPPAFERIEVLKGAASSSSFGAAPPGGSVNVITPTPSDR 201
      182 QGHAAGFGTHKQYKADADVSGSLNSDGYKRVMAQTVGASPPRAKNNRHETFYAAD 241
      202 FGEVYGTGSGFSHKEIGFDGDNLTADETL-----SYRLTKFQRSDAEIYISQ 250
      242 WDN-----PDVTLGAGYLYQORHLAPYNGLPADANNKL-PSLPQHVFGAD--W 288
      251 DDENFWGVGTWRPTDASTLTFIFD--HL-DKQGVPGSGGHPILGTDPRDQFGEPPDYF 307
      289 NKRRNSHDVFAULKHVFGNGYGVKVMRYSDDRADSNVAFAGSKLGMKTPAGPCNTA 348
      308 SETNRNSYSLVFD--HDFGNGLSFSSNARYSNLNDGFSAYIGA-----TP 351
      349 DDKACAVG--LGTEIKOKALAEFASYSRPRFGNTANEVYIGADYRFRSTNEQGRITL 405
      352 TDGSTVAGRYFPGNEKSTDPFVIDAHLVYASLDNVESRTLPGADYKYESD----- 403
      406 YARGALANEFRSIPQVDLIANARKGVRSYHTVATEMLDE--FGIYKSTFHPADGLSL 463
      404 -----SANFYAPASIDWEDPIYSGGAPAPAYASTNNQQTNAIYLQDDLPFPDKLTV 457
      464 IGGG-----LGHYKIESGEGKTLHKASKTKFTGYAGAVDLDNNSLYSLSLQLYPOT 518
      458 SFLGRNDWLDLSETNLLAG--TRAGNHRFETTRIGASYKYTEELAPYSYAESAP-- 512
      519 NLDADEKLKPRQGNQFEVYKGSYMDRLNA--RVSYFYMKOKN-----AAPLNPNNK 571
      513 --PAAS--DPTGKQYEVGK--YRPDAPFAPFTASVYDLTKGNTVFPQVYTLPTQVE 566
      572 KTRYAALGKRVMEGVETESGAVTPKMQIHAGYSYLHSQIKTASNSRDEIPLMPKMSA 631
      567 KYRH-----RGFELEAKAEVTNNISVLAAYSIDSKIEEPGANDGRMLKRVPRMA 618
      632 NIMTYQVTPELLIGG--GVNA-----MSGITSSAGMHAGGYATPDMAAARFTPKLK 662
      619 SWGTYTLLEGDARGDMLFELGARYTDAYVTSITNTSSBSA--VWFDAAFTYKIOENTT 676
      683 LQINADINFN-RHYVARVGANTENIPGSERTWTANLRYSF 722
      677 POLNVNMLFDEKRVASKDSGAVYVN-PG-RSLATILRKSW 714

RESULT 42
O9EZ93 PRELIMINARY; PRT; 762 AA.
ID O9EZ93:
AC 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Ferrichrome-iron receptor (Fragment).
OS Zymomonas mobilis.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Zymomonas.
NCBI_TaxId=542;
RN NCBI_TaxId=542;
RP SEQUENCE FROM N.A.
RC STRAIN-ZM4.
RA Shin I.S., Kang H.S.;
RT "Zymomonas mobilis ZM4 fosmid clone 43E12 complete sequence.";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF300471; AAG42412.1; -.
DR HSSP; P06971; LOJO.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Receptor.
FT NON TER
SQ SEQUENCE 762 AA; 83046 MW; F489525A093B358E CRC64;
Query Match 10.9%; Score 410.5; DB 2; Length 762;

```

```

      32 DLEFVHIKGRSNATITEKNGDYSSPAVTVGKIPASLREIOSVLIINQYKDNVD 91
      41 EVKTPSKKKDKSDVIVVATG--LSQASSTTKHTP--IIESPOSISTVSARDEIELASP 97
      92 TFDOLARKTPGLRV-LSNDDGR--SSVYARGY--EYSEVN--IDG--LPAWOMOSINGTL 142
      98 TINDALAYTGVQAEPBGIDSRVDEVSVRFGAGCRSSNNNPFVDGLRLPGSGQWTR-TSF 156
      143 NLPAFDRVEVMRGPSSGLFSSGEMGIVNLVRRKPTKAFQGH--AAAGFTHK--QYK 196
      157 DPESLQOIEVLKPGSGALYQQTAPGGVNLVTRKPTKSGGEFFLQAGY--TLGAWMQG 215
      197 AEAADVSGSLNSDGSVGRVMAQTVGASPPRAEKNNRHET--PFAAD--WDINPDT-- 248
      216 ASGDVSEKLNKDGTLGRI-----VGLARYGDTQVNVHVTGRYVYSPSLTWEISPDTKWTV 271
      249 -----VLGAGYLYQORHLAPYNGLPADANNKLPSLPQHVFGADWNNKF 291
      272 LAQYORDQGGSTQFLPATGTLYASKGRHA-----NDANICEPK-----WDF 315
      292 KNSHDVFAULKHVFGNGYGVKVMRYSDDRADSNVAFAGSKLGMKTY--PAGRPGCNT 347
      316 NRQVWAGSFEHFNH-----FTLRNNTRYTYLNTLYRATVLSGDTLQCSAALAGCVA 371
      348 AD--DCAVAGLTELKOKALADASYSRPRFGNTANEVIGADYRFRSTNEQGRITL 405
      372 GETNVRAVEGRG--RSQGVATDTQLEGHVDTGAVRHVWLFGTDVFTYDWTDRD-- 424
      406 YARGALANEFRSIPQVDLIANARKGVRSYHTVATEMLDEFGIY-----YSHVATENLDFGIY 450
      425 LVYSSLVL-----PMLNIFDPVRSAGYGAALSPOVYITSTISQDN--GIYFQDQK 474
      451 -----GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGYAGAVYDLN 501
      475 FKRLRYTVGGROQWAKDDTLNRLTGKR-----YLTSHNAPFTWRAGAVYLPD 520
      502 DNNSLYLSLQLYPOT--NLDADEKLKPRQGNQFEVYKGSYMDRLNAEAVSYRKM 558
      521 NGLAPFSYAKSPQVPSDSTSLDTPKPTGGDYEAIGIRYQY-GKSIYITFGAYQILT 579
      559 DKVAAPALPNPNKTRRYAALGKRVMEGV--ETRESGAVTPKMQIHAGYSYLHSQIK-TLS 615
      580 QRIMTTP-DPNGALCSTICLVOTEIEGVPVGGLEBRATLPMNMAIIFGTNSDAVYTS 638
      616 NSRSD-GIFL-LMPKHSANIMWTYQVT--DELTTGGVNAMS--GITSSAGMHAGGY 666
      639 NTASQIGNYLPQAPKMMASLFLIDRIRHGFPTGIGIGGVRYTGHSGYDPTNNT-LSIPGY 697
      667 ATPDMAAARFTPK-----LKLQINADNI FNRHYVARVGANTENIPGSERTWTANLRY 720
      698 TLFDSFLRYDFANRPPRYKGLSLISIMRNIAKRYAVATCTAASACYV--GQGRSLTLRLGY 756
      721 SF 722
      757 SW 758

RESULT 43
O8XVB1 PRELIMINARY; PRT; 801 AA.
ID O8XVB1:
AC 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative HYDROXAMATE-type ferrisiderophore receptor signal peptide
DE protein.
GN RSC2920 OR RS00173.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.

```


NCBI_TaxID=305;
[1]
SEQUENCE FROM N.A.
STRAIN=GMI1000;
MEDLINE=21681879; PubMed=11823852;
Salandub M., Genin S., Artiguenave F., Gouzy J., Mangnot S.,
Arlat M., Billault A., Brothier P., Camus J.C., Cattolico L.,
Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
Siquier P., Thebaud P., Whalen M., Wincker P., Levy M.,
Weissenbach J., Boucher C.A.;
"Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL; AL646072; CAD16627.1; --
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR Complete proteome.
KW SEQUENCE 801 AA; 87071 MW; 5577D637FF9203EA CRC64;
[1]
Query Match 10.9%; Score 410; DB 16; Length 801;
Best Local Similarity 22.5%; Pred. No. 6e-16;
Matches 188; Conservative 115; Mismatches 341; Indels 190; Gaps 28;
QY 10 INMTATVLAALSS-----SVFA-AQTADLETVHKGORS-----YNAIVTE 50
DB 37 IDPAEPDLAALNTLARQAGTQVFSALTAGTAPHVQGHYSVEETLNLLAQSLTAR 96
QY 51 KNGD-----YSSFVAVTVGTPKIPASLREIPQ 75
DB 97 RGDNFTTIEPAGSAGVETLPATVTTAAQESANGPVHGVARRGATGKTDPATIEEIPQ 156
QY 76 SVSIITNQVKRNVDFTDQARKTFLGRVLSNDDGRSSV--ARGYEYSEYNIDGLPAQ 133
DB 157 SISVTRDRPEAQAQSVNEALRYTASVYAGT--RSDWYTVARGFVPATY-LDGL--- 211
QY 134 MQSINGTLPNL-----FADFVEMVRGSPGLFDSGEMGGIVNLVRKPTKAFQ 183
DB 212 -----QLPNTINLASWRVDPWQIECVLLRGSPASVLYGQDGGTVVVTQPTAEPT 265
QY 184 HAAAGFTTHQYKAEADVSGSLNSDGSVRGRVMAQTVGASPRAEKNRHETFYAADWD 243
DB 266 EIEVQYGVSHVRKQVAGDFGALDADGKLSRLTALARDGNLPLGPFNDQRLMATSITWK 325
QY 244 INPDVLG--AGVLYQORHLAPYNGLP-----DANNKLPSPQVFGA--DWNKFKMN 294
DB 326 PSADTTLTLLANVLRDKN--ASDSFLPAQGTILPNPNKGPATR---FTGSPDFDAYEKT 381
QY 295 SHDYFADLKHFGNGYGVKVMYRSDRSDANSYAFAGSKLGMKTPAGRPCNCTADDKACA 354
DB 382 QYSLGYQFHERLDDTWTVRQARYMHLRLDDRMVY--GVGLDPSDPTQRTLLRVAGTAP 440
QY 355 VGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGLALN 414
DB 441 YG-----RLDIDTQQAARFATGMVHRLLAGVDYQR-----QTT----- 474
QY 415 EFRSIPQVDLLIANARKVGRGYSHTVATENLDFGIYKSTFHPADGLSLTGG----- 467
DB 475 -----RDPENYALTSLDLY---NPVQPTDRSIFSPGSPQDIR 512
QY 468 ----RLGHY---KTESGE-----GKTLHKASKTKFTYAGAVYD 499
DB 513 QRQTGLGLYLQDLKGLGRWVFTLGRHDWADTRTDDLGGTTTASQORDAFTGRVGVYVL 572
QY 500 LNDNNSLYLSLSQLYTPQTNLDADCKLLKPRQGNQFVEGYK---GSYMDRLNARVSFYR 556
DB 573 APYGISPLYSYSTSNFVTGTRDNOQAPKPYKGRQVEAGVKVQAGS-----KASVTASVFR 628
QY 557 MKDKNAAPLNPNNKKTRYAALGKRMVEGETEISGAVTPKQWIGHAGSYLHQSITASN 616
DB 629 INQRNVLPDLDPACTHSTQTGEVRSQGFEEATVQPMAGNVNVIAGIYQDVNRNTRAND 688
QY 617 SRDD--GIFLLMPKHSANLWTTVQVTPELTITGGVNA--MSGITSSAG-----MHAGGYAT 668

Db 478 -----ADKEFVD-----KALAKGIFENNAIOR-----FNSLSDSSP 510
Qy 503 NNSLYLSLSQLYTPQTN-----LDADGKLKPRQGNQPEVYKGSYMDRLNAR- 551
Db 511 NKKATNRRSYVPLHTHTDDRMGKLDLTGYGLFGREHDFVGY--AYGDEKIRSEI 568
Qy 552 VSFYRKDKNAAPLNPNNKTRRYAALGKRWEGVETE-----SGAV 594
Db 569 LEIYERRR-----VRPNTGATGYYAGS--CGSRTVICLLMSGGI 609

RESULT 45

P72599 PRELIMINARY; PRT; 863 AA.
AC P72599;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Ferrichrome-iron receptor.
GN FH04 OR SL11409.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Aemitsu B., Nakamura Y.,
RA Miyajima N., Hirose M., Sugitani M., Saezaki S., Kimura T.,
RA Hoshino T., Matsuno A., Muraki A., Nakazaki N., Naito K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90899; BAA16599.1;
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 863 AA; 95026 MW; F8190445F9063CE4 CRC64;

Query Match 10.6%; Score 402; DB 16; Length 863;
Best Local Similarity 22.2%; Pred. No. 2e-15;
Matches 171; Conservative 127; Mismatches 335; Indels 136; Gaps 31;

Qy 9 RINMTA-ATVLA-----LSSVFAAQTADLETVNHIKGRSYNAIVT-EKNGDYSSF----- 58
Db 150 RIVTNGNGVTPAQAQVPSSENLISLTPPINTV--ESEEEIRIVATREBEAAVQEFFVFN 207
Qy 59 -AVTVGTIKIPASLREIPOSVITITNOQVDRVDTFDOLARKTPGLRYLSNDGRSSVYA 117
Db 208 TSVAGTDTLP--IMDTPSAQVSEEVIRSQQAITLEDEVLTWSSVTFGGTGGRETIIFG 265
Qy 118 -RGY--EYSEYNIDGLPAQOMOSIN-----GTLPNLFADRYEVMRGPGLPDSGEMGSI 169
Db 266 IRGFENQFS-----DTVPILRDGFRLYGGFQGITVSHLQVQVFLKGPSSILVQIEPGGV 321
Qy 170 VNLVXKRPRTAFQGHAAAGFGTHKQYKAADVSGSLNSDGSVGRVM-AQTVGASPRPAE 228
Db 322 INLSKKPLNEPFAVEVOLGNGQIVRRFRDISGGLNPSGNLRYRLNGVYSNEASFRDFN 381
Qy 229 KNNHEFTFAADWDINDPTVLGAGYLYQORHLAPYNGI-----PAD 270
Db 382 QPLERFAFAPITVTYAITDDTSLAVEYINDTNPADPGLSSFGDGAAPVPSRVINDPSD 441
Qy 271 ANNTKLPSLPQHFVFGADNNKFRMNSHDVPAFLKHYFGNGGYGKVGARYSDRDADSNYAF 330
Db 442 IVNK-----NFTSAGYN-----LEHRFNEWMKLRNAPRYMSYNDYVNV--- 479
Qy 331 GSKGKMTFPAKPGCNT--ADDKACAVGLGTRIKQKALAFDASYSRPRLTANTANEFVIG 388
Db 480 ---IALPTVNGPTVTRFPADODG-----QOGSYSFYTNVAVGKFTGSKVKEHLLAG 527

Qy 389 ADVNFRSTNEQRTLLYARGGLAN-----EPRSIPQVDLINARKGVGSHVATENU 444
Db 528 IDYN-----WSESIILTFPG-GPRTSIVFPDPDINALPKEN-----RDDLPLFGDTFSSN- 576
Qy 445 DEFGIYKSTFHPADGLSLIGGRGLGHYKIESGEGKT-LHKASKTKFTGYA-----GAVY 498
Db 577 -RLGIYLDQVSLLENLILVAGLRVDTITQNTNNLQTDENQGNTOQSDAVTPRIGLLY 635
Qy 499 DLNDNNSLYLSQLYTPQTNLDADGKLKPRQGNQPEVYKGSYMDRLNARVSFRNK 558
Db 636 RPIPEISFPNSYQSFTPNSGIDISGNPLEPBERGEGFEGVAAELFEQDLTLTYFNIS 695
Qy 559 DKNAAA--PLNPNKKTRRYAALGKRWEGVETEISGAVTPKQOIHAGSYLHSOI--KTA 614
Db 696 KNNVAVSDPVNPLFLST-----IGTQOSQSELDIYGEIILPKKIIQNTSYINAKVETD 751
Qy 615 SNRRDGLIFLLMPKHSANIMTYQVTPBELTIGGVNAMSGITSSAGM----- 662
Db 752 PNFVDNRLEFGI-PYNNANLMTTYEIQS-----GALQGLGFGIGFNVVGDPRFDLANT 802
Qy 663 --AGGATFTDAAAYRFTPKLQLQINADNIPNRHYARVGGANTRNIPG 709
Db 803 YTVGDYIIIGNAAIFYO-RDKYRVALLNRFTNANYVRAVSGNQTGLEPG 850

Search completed: December 18, 2002, 06:44:53
Job time : 62.0531 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 05:44:19 ; Search time 44.9653 Seconds
(without alignments)
2139.581 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776
Sequence: 1 MQGFMSVFRINMTATVLA.....NTFNIGSERTWTANLRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
```

Database :
A_Geneseq.101002.*

1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT *
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT *
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT *
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT *
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT *
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT *
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT *
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT *
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT *
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT *
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT *
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT *
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT *
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT *
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT *
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT *
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT *
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT *
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT *
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT *
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT *
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT *
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3776	100.0	722	21	AA07697	A. <i>Neisseria mening</i>
2	3592	99.1	691	21	AA07698	A. <i>Neisseria mening</i>
3	3244	85.9	708	21	AA75566	<i>Neisseria meningit</i>
4	821.5	21.8	815	22	AAU3610	<i>Pseudomonas aerugi</i>
5	784	20.8	725	20	AAV38631	<i>Neisseria meningit</i>
6	784	20.8	725	20	AAV38634	<i>Neisseria gonorrhoe</i>
7	778	20.6	725	20	AAV38632	<i>Neisseria meningit</i>
8	765	19.5	729	22	AAV7565	<i>Neisseria gonorrhoe</i>
9	737.5	18.5	729	22	AAU3433	E. coli cellular p
10	676.5	17.9	623	20	AAV38633	<i>Neisseria gonorrhoe</i>

11	397.5	10.6	783	21	AAV75533	Neisseria meningitidis
12	400.5	10.5	783	21	AAV75554	Neisseria meningitidis
13	375.5	9.9	702	22	AA568923	Neisseria meningitidis
14	374.5	9.9	702	22	AA568921	Neisseria meningitidis
15	371	9.8	663	22	AB552462	Escherichia coli p
16	370	9.8	703	22	AA568926	Neisseria meningitidis
17	370	9.8	703	21	AAV75568	Neisseria meningitidis
18	370	9.8	703	22	AA568916	Neisseria meningitidis
19	368	9.7	703	21	AAV75567	Neisseria meningitidis
20	368	9.7	703	22	AA568927	Neisseria meningitidis
21	367.5	9.7	702	22	AA568920	Neisseria meningitidis
22	366	9.7	703	21	AA518720	A. Neisseria meningitidis
23	364.5	9.7	702	22	AA568922	Neisseria meningitidis
24	363	9.6	703	22	AA568919	Neisseria meningitidis
25	316.5	8.4	697	22	AA568917	Neisseria meningitidis
26	316.5	8.4	697	22	AA568925	Neisseria meningitidis
27	313.5	8.1	721	22	AB552899	Escherichia coli p
28	307.5	8.1	700	22	AA568918	Neisseria meningitidis
29	294.5	7.8	696	17	AA575366	Adhesin, Escherichia coli
30	273.5	7.2	725	22	AAV72224	E. coli ironEc ext
31	241.5	6.4	697	22	AA568924	Neisseria meningitidis
32	237.5	6.3	811	22	AAU33580	Pseudomonas aeruginosa
33	235.5	6.2	211	20	AAV38830	Neisseria meningitidis
34	228	6.0	660	22	AB552669	Escherichia coli p
35	223.5	5.9	682	22	AB552895	Escherichia coli p
36	220	5.8	785	21	AA180086	TonB dependent receptor
37	213	5.6	188	21	AAV75536	Neisseria meningitidis
38	205	5.4	187	21	AAV75535	Neisseria meningitidis
39	205	5.4	774	22	AAU29336	Novel human regulate
40	196	5.2	454	22	ABG17860	Novel human diaphan
41	195	5.2	134	21	AA584099	Human ORF ORP673
42	195	5.2	134	21	ABP06575	Human ORF protein
43	195	5.2	340	22	ABG29752	Novel human diaphan
44	193.5	5.1	833	20	AA344444	Porphyrinomas gingiv
45	193.5	5.1	876	20	AAV34368	Porphyrinomas gingiv

ALIGNMENTS

	RESULT 1
XX	AAB07697
ID	AAB07697 standard; Protein; 722 AA.
XX	
AC	AAB07697;
XX	
DT	07-NOV-2000 (first entry)
XX	
DE	A Neisseria meningitidis BASB053 polypeptide.
XX	
KM	BASB053; Neisseria meningitidis infection; vaccine.
XX	
OS	Neisseria meningitidis.
XX	
PN	WO200042193-A1.
XX	
PD	20-JUL-2000.
XX	
FE	10-JAN-2000; 2000WO-EF00137.
XX	
PR	15-JAN-1999; 99GB-0000959.
PR	28-JAN-1999; 99GB-0001903.
XX	
PA	(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
XX	
EI	Ruelle J;
XX	
DR	WPI; 2000-476062/41.
DR	N-PSDB; AAA59216.
XX	
PT	New Neisseria meningitidis polypeptide useful for diagnosis of
PT	Neisseria infection and for development of vaccines against such
PT	infection -


```
XX PS Claim 3; Page 55-56; 92pp; English.
XX
CC The present sequence represents a Neisseria meningitidis BASB053
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
CC for BASB053, may be identified in a biological sample in order to
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
CC polypeptides and polynucleotides may be used as vaccines, for
CC generating an immune response in an animal. A composition comprising
CC at least one antibody immunospecific for BASB053 may be used to
CC treat humans infected with Neisseria meningitidis.
XX
SQ Sequence 722 AA;
Query Match 100.0%; Score 3776; DB 21; Length 722;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MGQFMSVFRINNTAATVLAALSSVFAAQTADLETWHIKGQSYNAIVTEKNGDYSSFAV 60
Db 1 MGQFMSVFRINNTAATVLAALSSVFAAQTADLETWHIKGQSYNAIVTEKNGDYSSFAV 60
Qy 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDOLARKTPGLRLVLSNDDGRSSVYARGY 120
Db 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDOLARKTPGLRLVLSNDDGRSSVYARGY 120
Qy 121 EYSEYNIDGLPAQMOSINGTLPNLPFAFDRVEVMRGPGLFSSGEMGGIVNLVRKRPTKA 180
Db 121 EYSEYNIDGLPAQMOSINGTLPNLPFAFDRVEVMRGPGLFSSGEMGGIVNLVRKRPTKA 180
Qy 181 FQCHAAAGFTGHQYKAEADVSGSLSDGSRVGRVMAQTVGASPRPAEKNNRHETFYAAA 240
Db 181 FQCHAAAGFTGHQYKAEADVSGSLSDGSRVGRVMAQTVGASPRPAEKNNRHETFYAAA 240
Qy 241 DWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLPSPLOHVFVGDADNKFKNNSHDVFA 300
Db 241 DWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLPSPLOHVFVGDADNKFKNNSHDVFA 300
Qy 301 DLKHVFGNGGYGKVGWRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTE 360
Db 301 DLKHVFGNGGYGKVGWRYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTE 360
Qy 361 IKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIP 420
Db 361 IKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIP 420
Qy 421 QVDLIANARKGVGYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGK 480
Db 421 QVDLIANARKGVGYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGK 480
Qy 481 TLHKASKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQFEVGYK 540
Db 481 TLHKASKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQFEVGYK 540
Qy 541 GSYMDRLNARVSFRMKDKNAAPLNNKTRAAALGKRWMEGVETEISGAVTPKQKI 600
Db 541 GSYMDRLNARVSFRMKDKNAAPLNNKTRAAALGKRWMEGVETEISGAVTPKQKI 600
Qy 601 HAGSYLHSOIKTASNSRDDGIPLMPKHSANLWTTYQVTPELTIGGGVNMAGTSSAG 660
Db 601 HAGSYLHSOIKTASNSRDDGIPLMPKHSANLWTTYQVTPELTIGGGVNMAGTSSAG 660
Qy 661 MHAGGYATFDMAAAYRFTPKLQKLNADNIPNRHYARVGGANTFNI PGSERTWTANLRY 720
Db 661 MHAGGYATFDMAAAYRFTPKLQKLNADNIPNRHYARVGGANTFNI PGSERTWTANLRY 720
Qy 721 SF 722
Db 721 SF 722
RESULT 2
AAB07698
```

```
ID AAB07698 standard; Protein; 691 AA.
XX
AC AAB07698;
XX
DT 07-NOV-2000 (first entry)
XX
DE A Neisseria meningitidis BASB053 polypeptide.
XX
KW BASB053; Neisseria meningitidis infection; vaccine.
XX
OS Neisseria meningitidis.
XX
PN WO2000042193-A1.
XX
PD 20-JUL-2000.
XX
PF 10-JAN-2000; 2000WO-EP00137.
XX
PR 15-JAN-1999; 99GB-0000959.
PR 28-JAN-1999; 99GB-0001903.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PI Ruelle J;
DR WPI; 2000-476062/41.
DR N-PSDB; AAA59217.
XX
PT New Neisseria meningitidis polypeptide useful for diagnosis of
PT Neisseria infection and for development of vaccines against such
PT infection -
XX
PS Claim 3; Page 56; 92pp; English.
XX
CC The present sequence represents a Neisseria meningitidis BASB053
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
CC for BASB053, may be identified in a biological sample in order to
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
CC polypeptides and polynucleotides may be used as vaccines, for
CC generating an immune response in an animal. A composition comprising
CC at least one antibody immunospecific for BASB053 may be used to
CC treat humans infected with Neisseria meningitidis.
XX
SQ Sequence 691 AA;
Query Match 95.1%; Score 3592; DB 21; Length 691;
Best Local Similarity 99.7%; Pred. No. 3.8e-318;
Matches 684; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 37 HIKQORSYNAIVTEKNGDYSSFAVTVGTGKIPASLRIPQSVSIITNQOVKDRNVDTFDQL 96
Db 6 HIKQORSYNAIVTEKNGDYSSFAVTVGTGKIPASLRIPQSVSIITNQOVKDRNVDTFDQL 65
Qy 97 ARKTPGLRVLNDDGRSSVARGVEYSEYNIDGLPAQMOSINGTLPNLPFAFDRVEMRGP 156
Db 66 ARKTPGLRVLNDDGRSSVARGVEYSEYNIDGLPAQMOSINGTLPNLPFAFDRVEMRGP 125
Qy 157 SGLFDDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSGSLSDGSRVGRVM 216
Db 126 SGLFDDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSGSLSDGSRVGRVM 185
Qy 217 AQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLP 276
Db 186 AQTVGASPRPAEKNNRHETFYAAADWDINPDTVLGAGLYLQORHLAPYNGLPADANNKLP 245
Qy 277 SLPOHVFVGADWNKFKWNSHDVFDADLKHVFGNGGYGKVGWRYSDRDADSNYAFAGSKLGM 336
Db 246 SLPOHVFVGADWNKFKWNSHDVFDADLKHVFGNGGYGKVGWRYSDRDADSNYAFAGSKLGM 305
Qy 337 KTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRS 396
Db 306 KTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRS 365
```


QY 397 TNEQGGTTLVARGGLALNFRSIPQVDLIANARKVGRGSHVATENLDEFGIYKSTPH 456
 Db 366 TEEQGGTTLVARGGLALNFRSIPQVDLIANARKVGRGSHVATENLDEFGIYKSTPH 425
 QY 457 PADGLSLIGGRLGHYKISSEGGKTLHKASKTFGTGAGAVVDLNNNSLYLSLSQLYTP 516
 Db 426 PADGLSLIGGRLGHYKISSEGGKTLHKASKTFGTGAGAVVDLNNNSLYLSLSQLYTP 485
 QY 517 QTNLDADGKLLKPRQGNQEVGKGSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYA 576
 Db 486 QTNLDADGKLLKPRQGNQEVGKGSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYA 545
 QY 577 ALGKRVMEGVETEISGAVTPKWOIHAGYSYLSQITASNSRDDGIFLLMPKKSANLMTT 636
 Db 546 ALGKRVMEGVETEISGAVTPKWOIHAGYSYLSQITASNSRDDGIFLLMPKKSANLMTT 605
 QY 637 YQVTPPELTIGGGVNAMSGITSSAGMAGGATPDMAAAYRFPKLTQINADNIFNRHY 696
 Db 606 YQVTPPELTIGGGVNAMSGITSSAGMAGGATPDMAAAYRFPKLTQINADNIFNRHY 665
 QY 697 ARVGANTFNIPGSERTWTANLRYSF 722
 Db 666 ARVGANTFNIPGSERTWTANLRYSF 691

RESULT 3
 AAY75566
 ID AAY75566 standard; Protein; 708 AA.
 XX
 AC AAY75566;

DT 21-MAR-2000 (first entry)
 DE Neisseria meningitidis ORF 760 protein sequence SEQ ID NO:2606.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KM antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.

XX Neisseria meningitidis.

OS
 XX WO9597280-A2.

PN 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

PF 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (GENO-) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Masigman V, Mora M,
 PI Petersen J, Pizzo M, Rappoli R, Ratti G, Scalato E, Scarselli M,
 PI Tettelin H, Venter JC;

XX MPI: 2000-062150/05.

DR N-PSDB; AA254328.

XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics

PS Claim 2; Page 1235; 1453bp; English.

XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC Neisseria bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of Neisseria bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 708 AA;

Query Match 85.9%; Score 3244; DB 21; Length 708;

Best Local Similarity 86.8%; Pred. No. 2,5e-286;
 Matches 627; Conservative 33; Mismatches 48; Indels 14; Gaps 3;

QY 1 MQQFMSVFRIINNTAATVLAALSSVFAAQTADLETVHIKGRSYNAIVTERKNDYSSFAV 60
 Db 1 MQQFMSVFRIINNTAATVLAALSSVFAAQTADLETVHIKGRSYNAIVTERKNDYSSFAV 60
 QY 61 TVGTIKIPASLRBIPOSVSLITNQVADRVDFTDQARXTPGLRVLNSDGRSSVYARGY 120
 Db 61 TVGTIKIPASLRBIPOSVSLITNQVADRVDFTDQARXTPGLRVLNSDGRSSVYARGY 120
 QY 121 EYSEYVNDGLPAWOSINGTLPLNLFADREVEYMRGSGLFDSSGEMGIVNLVRRKPTA 180
 Db 121 EYSEYVNDGLPAWOSINGTLPLNLFADREVEYMRGSGLFDSSGEMGIVNLVRRKPTA 180
 QY 181 FQGHAAAGFTHKQYKAADVSGSINSDSVGRVMAQTGASPRPAEKNNHEFTYAAA 240
 Db 181 FQGHAAAGFTHKQYKAADVSGSINSDSVGRVMAQTGASPRPAEKNNHEFTYAAA 240
 QY 241 DWDINPDTVLGAGYLYQOCHLAPYNGLPADANKLPSPQHVFVGDAMKPFKMSHDVPA 300
 Db 241 DWDINPDTVLGAGYLYQOCHLAPYNGLPADANKLPSPQHVFVGDAMKPFKMSHDVPA 300
 QY 301 DLKHYFNGGKGYKGRYSDDRADSNVAFAGSKLGWKTAPGRPGCNTADKCAVGLGTE 360
 Db 301 DLKHYFNGGKGYKGRYSDDRADSNVAFAGSKLGWKTAPGRPGCNTADKCAVGLGTE 360
 QY 361 IKQALAFDASYSRPRLGNTANEFVIGADYVNRFRSTNEQGRFTLLVARGGLALNFRSIP 420
 Db 349 IKQAFADVASYSRPALGNTANEFVIGADYVNRFRSTNEQGRFTLLVARGGLALNFRSIP 406
 QY 421 QVDLIANARKVGRGSHVATENLDEFGIYKSTHPADGLSLIGGRLGHYKISSEGGK 480
 Db 407 YNGIIONARAGKGFHVSYTEENLDETGLYAKTVRFRPLEGLSLIAGRGVGHKIESGDGK 466
 QY 481 TLHKASKTFGTGAGAVVDLNNNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQEVGK 540
 Db 467 TLHKASKTFGTGAGAVVDLNNNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQEVGK 526
 QY 541 GSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYAALGKRVMEGVETEISGAVTPKMOI 600
 Db 527 GSYMDRLNARVSFYRMKDKNAAAPLNNKKTTRYAALGKRVMEGVETEISGAVTPKMOI 586
 QY 601 HAGYSYLSQITASNSRDDGIFLLMPKKSANLMTTYQVTPPELTIGGGVNAMSGITSSAG 660
 Db 587 HAGYSYLSQITASNSRDDGIFLLMPKKSANLMTTYQVTPPELTIGGGVNAMSGITSSAG 646
 QY 661 MHAGGYATPDMAAAYRFPKLTQINADNIFNRHYARVGANTFNIPGSERTWTANLRY 720
 Db 647 IHAGGYATPDMAAAYRFPKLTQINADNIFNRHYARVGANTFNIPGSERTWTANLRY 706
 QY 721 SF 722
 Db 707 SF 708

RESULT 4
 AAU33610


```

QY 115 VYARGYEYSEYNI DGLPAQMSQINGTLPNLPAFDRVEMRGPSGLFDSSGEMGIVNLVR 174
DB 116 LFARGSRANYQINGIPVADALADTGNANTAAAYERVEVVRGAGLPDGTGSPSATVNLVR 175
QY 175 KPTTAFQCHAAAGFTGKQYKAEADVSGSLNSDGSVRCRVMAQ-TVGASPRPAEKNNRH 233
DB 176 KPTTRKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGRGDSWQLER-SRD 234
QY 234 ETFYAAADWDINPDTVLGAGYLYQQRHL---APYNGLPADANNKLPSPHQHVFVGADWNK 290
DB 235 AELYGILEYDIAPQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWSN 294
QY 291 FQNSHDVPADLKHVFGNGGKVGMRYSDBDADSNYAFAGSKLGMKTPAGPCGNTADD 350
DB 295 SRNRALNLFAGIEHRFN-----QDWKKAEDYDTRSR--FRQPYGVAGVLSIDH 341
QY 351 KACAVGL---GTEIKOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEOQRTTLA 407
DB 342 STAATDLIPGYWHADPRTHSASMSLTGKRLFGREHDLIAGINGKYAS-NKYGERSIIP 400
QY 408 RGLALNEF-----RSIPQVDLIANARKGVRGYSHTVATENLDEFIYGKST 454
DB 401 NAI PNAYEFSRTGAYPQPSSFAQTIPQYD---TRRQIGGYLAT-----R 441
QY 455 FHPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLDNNSLYLSLQ 512
DB 442 FRAADNLSLILGGYSRYRTGSDYSTRSQGMYTVSANRFTPYTGIVFDLTGNLSLYGSYSS 501
QY 513 LYTPTQNLADGKLLKPRQGNQFVGYKGSYMDRLNARVSFYRMKDN--AAAPLNPNK 571
DB 502 LFPVQKQDEHGSYLKPVGTNNLEADIKGEWLEGRNLASAAYRARKNNLATAAGRPDSG 561
QY 572 KTRYAALGRKVMGVETEISGAVTPKWKQIHAGYSYLHSQIKTASNRDDGIFLLMPKHS 631
DB 562 NTYVRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKPRDQGSRLNPDV---PERSF 617
QY 632 NLWTTYQVTPEL---TIGGVV-----NAMSGITSSAGMHAGGYATFDD 671
DB 618 KLFETAYHLAEPASGRTIGAGVRRQGETHTDPAALRIPNPAKARAVANSROKAYAVADI 677
QY 672 MAAYRFTPKLQIINADNIFNRHYARVGGANTFNIP-----GSERTWTANLRSYF 722
DB 678 MARYRPNRTELNLNDVNFNKHRTQ-----PDRHSYGALRTVNAATYRF 724

RESULT 7
ID AAY38832 standard; Protein; 725 AA.
XX AC AAY38832;
XX DT 08-OCT-1999 (first entry)
XX DE Neisseria meningitidis strain A antigen encoded by ORF23.
XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX OS Neisseria meningitidis.
XX PN WO9924578-A2.
XX PD 20-MAY-1999.
XX PF 09-OCT-1998; 98WO-IB01665.
XX PR 01-SEP-1998; 98GB-0019016.
XX PR 06-NOV-1997; 97GB-0023516.
XX PR 14-NOV-1997; 97GB-0024190.
XX PR 18-NOV-1997; 97GB-0024386.
XX PR 27-NOV-1997; 97GB-0025158.
XX PR 10-DEC-1997; 97GB-0026147.
XX PR 14-JAN-1998; 98GB-0000759.

```

```

XX (CHIR-) CHIRON SPA.
XX Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
XX WPI; 1999-327407/27.
XX N-PSDB; AAZ12260.
XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
XX diagnosis, treatment and prevention of infection
XX Claim 4; Page 381; 524pp; English.
XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
XX and N. gonorrhoeae antigenic proteins. They are encoded by open
XX reading frames (ORFs) AAZ11972-212358. The antigenic proteins,
XX their fragments, their nucleic acids and antibodies are used for
XX diagnosis, prevention (as vaccines) or treatment of Neisseria
XX infections, such as meningitis, septicaemia and gonorrhea. Both
XX organisms are closely related. Fragments of the nucleic acids
XX are useful as hybridisation probes and antisense reagents.
XX SQ Sequence 725 AA;
XX Query Match 20.6%; Score 778; DB 20; Length 725;
XX Best Local Similarity 29.7%; Pred. No. 1.6e-61;
XX Matches 231; Conservative 117; Mismatches 317; Indels 112; Gaps 23;
QY 5 MSVFRINWTAATVL-----AALSSSVFAAQTADLETVHIKGORSYNAIVTEKNGDYSS 57
DB 1 MTRFKYSLFPAALLPVVAQADVSDDPKQESTELTITVADRT-----ASNDGYTV 55
QY 58 FAVTVGTKIPASLREIPOSVSIITNQVKDRNVDTFQ-LARKTPGLRVLNSND--GRSS 114
DB 56 SGTHTPLGLPMTLEIPOSVSVITSQMRQDNKALDRALLOATGTSRQIYSGDRAGNY 115
QY 115 VYARGYEYSEYNI DGLPAQMSQINGTLPNLPAFDRVEMRGPSGLFDSSGEMGIVNLVR 174
DB 116 LFARGSRANYQINGIPVADALADTGNANTAAAYERVEVVRGAGLPDGTGSPSATVNLVR 175
QY 175 KRPTKAFQGHAAAGFTGKQYKAEADVSGSLNSDGSVRCRVMAQ-TVGASPRPAEKNNRH 233
DB 176 KRPTRKPLFEVRAEAGNRKHFGLGADVSGSLNAEGTLRGLVSTFGRGDSWQLER-SRD 234
QY 234 ETFYAAADWDINPDTVLGAGYLYQQRHL---APYNGLPADANNKLPSPHQHVFVGADWNK 290
DB 235 AELYGILEYDIAPQTRVHAGMDYQAKETADAPLSYAVYDSQGYATAFGPKDNPATNWSN 294
QY 291 FQNSHDVPADLKHVFGNGGKVGMRYSDBDADSNYAFAGSKLGMKTPAGPCGNTADD 350
DB 295 SRNRALNLFAGIEHRFN-----QDWKKAEDYDTRSR--FRQPYGVAGVLSIDH 341
QY 351 KACAVGL---GTEIKOKALAFDASYSRPPRLGNTANEFVIGADYNRFRSTNEOQRTTLA 407
DB 342 STAATDLIPGYWHADPRTHSASMSLTGKRLFGREHDLIAGINGKYAS-NKYGERSIIP 400
QY 408 RGLALNEF-----RSIPQVDLIANARKGVRGYSHTVATENLDEFIYGKST 454
DB 396 -----RSI-----IPNAIPNAYEFSRTGAYPQPSSFAQTIPQYGTTRQIGGYLATR 441
QY 455 FHPADGLSLIGGRLGHYKIESGEGKT--LHKASKTKFTGYAGAVYDLDNNSLYLSLQ 512
DB 442 FRAADNLSLILGGYSRYRTGSDYSTRSQGMYTVSANRFTPYTGIVFDLTGNLSLYGSYSS 501
QY 513 LYTPTQNLADGKLLKPRQGNQFVGYKGSYMDRLNARVSFYRMKDN--AAAPLNPNK 570
DB 502 LFPVQKQDEHGSYLKPVGTNNLEADIKGEWLEGRNLASAAYRARKNNLATAAGRPDSG 561
QY 571 KTRYAALGRKVMGVETEISGAVTPKWKQIHAGYSYLHSQIKTASNRDDGIFLLMPKHS 630
DB 562 NTYVRAANQAKTHGWEIEVGGRIITPEWQIQAGYSQSKTRDQGSRLNPDV---PERS 616
QY 631 ANLWTTYQVTPEL---TIGGVV-----NAMSGITSSAGMHAGGYATFD 670

```



```
Db 617 FKLTYYHAPAEPSGWTIGAGVFMOSSETTTPATLRIPPAKAPADNSROKAAVAD 676
Qy 671 AMAAYFTPEKLKQINADNIFNNHYARVAGANTFNI-----GSERTWTANLRYSF 722
Db 677 IMARVRFNRAELSLVNDLPNHGHTQTQ-----PDHSGALRTVNAATYRF 724

RESULT 8
AAV75565
ID AAV75565 standard; Protein; 154 AA.
XX
AC AAV75565;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 760 protein sequence SEQ ID NO:2604.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9957280-A2.
XX
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WO-US09346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-00989062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
PA (CHIR) CHIRON CORP.
XX
PA (GENO-) INST GENOMIC RES.
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M,
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M,
PI Tettelein H, Venter JC;
XX
DR WPI: 2000-062150/05.
DR N-PSDB; AA254327.
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1234; 1453pp; English.
XX
CC AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenetic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 154 AA;
XX
Query Match 20.3%; Score 765; DB: 21; Length 154;
Beet Local Similarity 92.8%; Pred. No. 2e-61;
Matches 143; Conservative 6; Mismatches 5; Indels 0; Gaps 0;
```

```
Qy 569 NNKRTYALGRVMEGVETEISGAVTPKQIHAGSYLHSGIKTASNSRDDGIFLLMPK 628
Db 1 NNRNTRYALGRVMEGVETEISGAVTPKQIHAGSYLHSGIKTANPRDDGIFLLVPR 60
Qy 629 HSNALMTTYQVTPBELTIGGVNAMSIGTSSAGMHAGYATFPDMAAYRTPKQKQIND 688
Db 61 HSNALMTTYQVTPGLTVGGVNAMSGITSSAGMHAGYATFPDMAAYRTPKQKQIND 120

Qy 689 NIFNRHYARVAGANTFNI-----GSERTWTANLRYSF 722
Db 121 NIFNRHYARVAGANTFNI-----GSERTWTANLRYSF 154

RESULT 9
AAU34539
ID AAU34539 standard; Protein; 729 AA.
XX
AC AAU34539;
XX
DT 14-FEB-2002 (first entry)
XX
DE E. coli cellular proliferation protein #120.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PR 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI: 2001-611495/70.
DR N-PSDB; AAS52398.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 10132; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
```


CC format directly from WIFO at
CC ftp.wifo.int/pub/published_pct_sequences.

XX Sequence 729 AA;

Query Match 19.5%; Score 737.5; DB 22; Length 729;
Best Local Similarity 27.3%; Pred. No. 7.8e-58;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;

QY 3 QFMSVFRINMTATV-LAALSSSVFAAQADLETVHIKQGRSYNAIVTEKNGDYSFAVT 61
DB 11 QYCAITKPSLLAGCIALALLPSAFAAPE-ETVIVEGSAT---APDGDENDYSVTS 66
QY 62 VGTKIPASIREIPQSIIITNQVKDRVDTFDQARKTPGLRVLSNDRGSSVARGVE 121
DB 67 AGTKMQMTQORDIPQSVTIYSQQRMDDQQLTGLGVMENTLGISKQADSRLALYSRGR 126
QY 122 YSEYNIDGLPAQMOS---INGTLPNLFAFDRVEMRGPGLFDSSGEMGGIVNLVKRPT 178
DB 127 IDNYWVDGIPTFESRWNLGDALSDMALFERVEVVRGATGLMTGTGNPSAANMVRKHAT 186
QY 179 -KAFQGHAAAGFGTHKQYKAEADVSGSLNSDGSVRGRVMAQTGVA---SPRPAEKNNRHE 234
DB 187 SREFKGVSAEYGSWKNERYVADLQSPLETDGKIRARI---VGGYQNDSWLDVYNSEK 242
QY 235 TTYAA-ADWDINDPDTVLGAGLYQORHL--APYNGLPA-DANNKLPSPQHVFGADWNK 290
DB 243 TPFSGIVADLGLDITLTSAGYEQRIDVNSPTWGLPRWNTDGSNSYDRARSTAPDWAY 302
QY 291 FKMNSHDVADLKHVFGNGYGVGMRYSDRDADSNY---AFAGSKLGMKTP----- 339
DB 303 NDKINKVFWTLKQQPADTQWATLNATHSEVFDSKQMYVDVAVNKADGMLVGPVSNYGP 362
QY 340 ----AGRPGCNTADKACAVGLGTE-----IKQALAFDASYSRPFR--LGNATNEF-- 385
DB 363 GFDYVGGTGWNSKRVKVDLDFADGSYELFGRQHNLMFGGSYKQNNRYFSSWANI 422
QY 386 VTGADYNRFRSTNEQRTLLYARGGLALNEFRSIPQVDLIANARKGVRSYSHVTATENLD 445
DB 423 EIGSFYN-FNG-----NFPQTD-----WSPQSLAQDDTT 450
QY 446 EF-GIYKSTFFPADGLSLIGGRGLHYKIESGEGKTLHKASKTKFTGYAGAVYDLNDNN 504
DB 451 HKSLFAATRVTLADPLHLILGARYTNMRVDT-----LTSMEKNHTTPYAGLVFDINDNW 506
QY 505 SLYLSLSQLYTPTQNLADGKLLKQGNQFQFVGYKGSYMDRLNARVSFYRMKDKNAA- 563
DB 507 STYASVTSIFQPNDRDSSGKYLAPITGNVYELGLKSDMNSRLITTLAIFRIEQDNVAQ 566
QY 564 ----APLNPNNKTRYAALGRVMEGVETEISGAVTPKQI HAGYS-YLHSQIKTASGRD 619
DB 567 STCTPIPSNGETAYKAVDGTGSKGVEFELNGAITDNWLTFCATRYI-----AEDNEG 620
QY 620 DGIFFLLMPKHSANLWTTYQ--VTPELTICGGYNAMSGITSSA-----GMHAGYATFD 670
DB 621 NAVNENLPPTVMKFTSYRLPMPELTVGGVWQNRVYTDVTVPYGFTRAEQGSYALVD 680
QY 671 AMAAARFTPKLKLQINADNIFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
DB 681 LFTRVQVTKNFSLOGVNVLFDKTYDNTVEGSIVY---GTPNFSITGYQF 729

RESULT 10

AAV38833

ID AAV38833 standard; Protein; 623 AA.

XX AC AAV38833;

XX DT 08-OCT-1999 (first entry)

XX DE Neisseria gonorrhoeae antigen encoded by partial ORF23.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
Neisseria gonorrhoeae.

WO9924578-A2.

20-MAY-1999.

09-OCT-1998; 98WO-IB01665.

01-SEP-1998; 98GB-0019016.

06-NOV-1997; 97GB-0023516.

14-NOV-1997; 97GB-0024190.

18-NOV-1997; 97GB-0024386.

27-NOV-1997; 97GB-0025158.

10-DEC-1997; 97GB-0026147.

14-JAN-1998; 98GB-0000759.

(CHIR-) CHIRON SPA.

Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;

WPI; 1999-327407/27.

P-PSDB; AAY38833.

Proteins from Neisseria meningitidis and N. gonorrhoeae useful for

diagnosis, treatment and prevention of infection

Claim 4; Page 383; 524pp; English.

Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis

and N. gonorrhoeae antigenic proteins. They are encoded by open

reading frames (ORFs) AAZ11972-212358. The antigenic proteins,

their fragments, their nucleic acids and antibodies are used for

diagnosis, prevention (as vaccines) or treatment of Neisseria

infections, such as meningitis, septicaemia and gonorrhea. Both

organisms are closely related. Fragments of the nucleic acids

are useful as hybridisation probes and antisense reagents.

Sequence 623 AA;

Query Match 17.9%; Score 676.5; DB 20; Length 623;

Best Local Similarity 30.0%; Pred. No. 2.2e-52;

Matches 198; Conservative 90; Mismatches 277; Indels 95; Gaps 17;

QY 111 GRSSVYARGYSEYNTDGLPAQMOSINGTLPLNLFADRVEMRGPGLFDSSGEMGGIV 170

DB 10 GYNYLFARGSRANYQINGIPVADALADTGNANTAAVERVEVRGVLPGDTGEP SATV 69

QY 171 NLVRKRPKAFQGHAAAGFGTHKQYKAEADVSGSLNSDGSVRGRVMAQ--TVGASPRPAEK 229

DB 70 NLVRKHPTKPLFEVRAEAGNRKHFGLCADVSGSLNAEGTLRLGLVSTFGRGDSWRQLER 129

QY 230 NNRHETFFAAADWDINDPDTVLGAGVLYQORHL---APYNGLPADANNKLPSPQHVFGA 286

DB 130 -SRDAELYGILEYDIAQTRVHAGMDYQQAETADAPLSYAVYDSQGYATAFGPKONPAT 188

QY 287 DWNKFKMNSHDVDFADLKHVFGNGYGVGMRYSDRDADSNYAFAGSKLGMKTPAGRPCCN 346

DB 189 NWSNSRNPALNLFAGIEHRFN-----QDKWLKAEYDYTSR--FPQPYGVAGVL 235

QY 347 TADKACAVGL---GTEIKQKALAFDASYSRPRLGNANTANEFVIGADYNRFRSTNEQORT 403

DB 236 SIDHSTAATDLIPGYWHADPRTHSASMSLTGKRLFGREHDLIAGINGYKYAS-NKYGER 294

QY 404 TLYARGGLALNEF-----RSIPOVDLIANARKGVRSYSHVTATENLDEFGIY 450

DB 295 SIINAPLNAYEFSRTGAYPOPSSFAQTIPQYD----TRRQIGGYLAT----- 338

QY 451 GKSTFFHPADGLSLIGGRGLHYKIESGEGKT--LHKASKTKFTGYAGAVYDLNDNNSLYL 508

DB 339 ---RFRADNLSSLILGGRYSYRAGSYNSRTQGTYYVSANRFTPTGTGIVFDLTGNLSLYG 395

QY 509 SISQLYTPQTNLDADGKLLKPRQGNQPEVGYKGSYMDRLNARVYPRMKDKY-AAAPLN 567
DB 396 SYSLFVLPOLQKDEHSGSYLKPVGNNLEADIKEMLEGRINASAAVYRRARKNNLATPAQR 455
QY 568 PNNKTRVYALGRVWEGVETESGAVTPKMOIHAGSYLHSGIKTASNRDDGIFLLMP 627
DB 456 DQSGNTYVPAANOAKTHGWEIEVGRITPEMOIQAGYSQSKPRDQGSRLNPDV---P 511
QY 628 KISANLMTTYQVTPEL---TIGGVY---NMSGITSSAGMHAGGYA 667
DB 512 EHSFPLFTYVHLAEPAPSGRTTIGAGVRQGETHTDPAALRIPPAKARAVANSRQKAYA 571
QY 668 TPDMAAAYFTPTPLKQIQINADNIFNRHYTARVCGANTFNIP---GSERTWTANLRYSF 722
DB 572 VADIMARVFRFPRTSLNVNDLFNHGYRQ---PDRHSYGALRTVAAAFYRF 622

RESULT 11
ID AAY75533 standard; Protein; 783 AA.
AC AAY75533;
XX
XX
XX 21-MAR-2000 (first entry)
DE Neisseria meningitidis ORF 742 protein sequence SEQ ID NO:2540.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KM antibacterial; gene therapy.
XX
XX Neisseria meningitidis.
OS
XX
XX WO957280-A2.
PN
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999; 99WO-US09346.
PF
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098894.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
PR
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzza M, Rappuoli R, Ratti G, Scalato E, Scarfelli M;
PI Tettelin H, Venter JC;
PI
DR WPI; 2000-062150/05.
DR N-PSDB; AA254295.
XX
XX
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
PS Claim 2; Page 1208; 1453p; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941,
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also

CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.

XX
XX
SQ Sequence 783 AA;

Query Match 10.6%; Score 400.5; DB 21; Length 783;

Best Local Similarity 21.4%; Pred. No. 4.6e-27;
Matches 170; Conservative 89; Mismatches 212; Indels 325; Gaps 28;

QY 237 YAAADWDINDPVDYLAGYLYQO-----RHLLPYNGLPADANNKLPSLPQ 280
DB 3 YGIAEADADSDVLTIGGYQKREVPDFSGIILPCENQKTAFFSSTP--ACNRPLQPR 60
QY 281 HVEVGADNMKFKMNSHDVADAKHYFGNG----- 309
DB 61 NTYLGEDWRLSADKYNLFSGRKHVFDNQMQLABVSYTKNESDAKVQFPLKNEVLA 120
QY 310 -GYKRYGM-----RYSRDAD-----SNYAF----- 329
DB 121 SGEDAVGFLTEKNEVLPPEPKDALEKLKAYRDETAKEYRERKDPFVKRPNNTAFEQYR 180
QY 330 -----AGSKLGKMTPA-----GRPGCNTADOKACAVGLGTIHXKALA----- 367
DB 181 SRRPAERKAGFDPKCMSDPPALDFTCGSWGDPGVDA--DKA-----EFVDKALKEGI 231
QY 368 -----FDASYSR-----PPR-----LGNANEF 385
DB 232 FNNAQRFNSLYDSSFNRKATANRRYSYMPLRHKDQDKGDKLDLTCTYGLFGREHNF 291
QY 386 VIGADY-----NFR----- 395
DB 292 FVGAYGDEKIRSEYLEIYERRVAVPNTGATHGVYAGSCQEPDGLSSPLVRGHEPD 351
QY 396 --STNEQRTLLYARGLANEFRSIPQ-----VDLIANARG 431
DB 352 WOAYDEKMRVTYVAEBCRNAKKIKTEPKLDABGQVYVYDEYSGSRTPVYVDVLEDEG 411
QY 432 VR-----QVSHT-----VATENLDEFGIYK-----STFHP 457
DB 412 NKIQETNPQGTAFGFSGITVPEWKTAVKADHDVAPALNYAKYALNTNKTSLTASRFV 471
QY 458 ADGLSLIGGRLGHYKIESGEKTL-----HKASKT---KFTG 492
DB 472 TGRHLHGG--LHYTRYETSGTKDMPVRVYQSPASDFQTASIRADQDHYTAKMQGHKLP 529
QY 493 YGCAVVDLNDNNSLYLSQLYTPQTNLDADGK-LIKPRQGNQPEVGYKGSYMDRLNAR 551
DB 530 YAGITVDLTPQOSIYGSYTKIFKQODNDVDSAKTVLPVGTNYEGWKGAPLQGRINAS 589
QY 552 VSFPRMKDKN-----AAAPLNPNKKTRYAALGKRVMEGVETESGAVTPKMOIHA 602
DB 590 FALFYLEQKNRVTVDGFGYVPAAGAGKQGSFOTYAKPIGKAVSGAEFELSGELNEDMKVA 649
QY 603 GYSYLHSGIKTAS-----NSRDD--GIFLMPKHSANLMTTYQV--TPBELTIGGYN 650
DB 650 GYTYNKSRYKNAEVAERLANSSADPVNFSNFPVHIFRPTGSHIINTGLTVGGVVS 709
QY 651 AMSGITSSAGMHAGGYATPDMAAAYFTPTPLKQIQINADNIFNRHY---YARVCGANTN 706
DB 710 AOSGTSILYINROGGYGLDGFVRVELGHAKLSLIGTNLNGRTYFENNYYNRTGRANPEY 769
QY 707 IPGSERTWTANLRYSF 722
DB 770 --GEPRVSMKLDWQF 783

RESULT 12
ID AAY75534 standard; Protein; 783 AA.
AC AAY75534;
XX
XX

DT 21-MAR-2000 (first entry)
XX Neisseria meningitidis ORF 742 protein sequence SEQ ID NO:2542.
DE
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
OS Neisseria meningitidis.
XX
PN WO9957280-A2.
XX
XX 11-NOV-1999.
PD
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
XX WPI; 2000-062150/05.
DR N-PSDB; AA254296.
DR
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 2; Page 1209; 1453pp; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AA274253 to AA275941
CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides
CC and polypeptides. AA254537 to AA254576 and AA254616 to AA254673 represent
CC PCR primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
SQ Sequence 783 AA;

Query Match 10.5%; Score 397.5; DB 21; Length 783;
Best Local Similarity 21.3%; Pred. No. 8.6e-27;
Matches 168; Conservative 92; Mismatches 220; Indels 307; Gaps 27;

QY 237 YAAADWDNDPTVLGAGVLYQQ-----RHLPYNGLPADANNKLPSPLPQ 280
DB 3 YGIAEADAGDSVLTGGNYQKRSVPDPFSGIILSCENQKTAFFSTP--ACNRPLQLPR 60
QY 281 HVFVGADWNKFKMNSHDVADLKHVFGNG-----GYGKVGMY----- 318
DB 61 NYILGEDWSRLSADKYNLPSFGFKHVDNGWQLNAEVSYTKNESDAKVQOFFLKNHHAAGL 120
QY 319 SDRDA-----DSNYAFA--G 331
DB 121 SDEDAVGFLTEKNEVIPPPEPKDALEKLKAYRDETAKEYRERKDDFVKNRPDNTAFEQYR 180
QY 332 SKLGMKTPAGRGCCNTAD---DKACAVGLG-----TEIKQKALA----- 367

Db 191 SRRAERKAGFDECMASAPFALDFICQSGWDPGVVDADKSEFVDKALAKEGIFNNAORFP 240
QY 368 ---FDASYR-----PFR-----LCNTANEFVIGADY--N 392
Db 241 NSLYDSSFNRRKATARRRYSYMPLRHTKDDRWGIGIKLDLTGYTGLFGREHDFVGYAYGDE 300
QY 393 RFRST-----NEQGR 402
Db 301 KIRSEYLEIYERRHRVRPNTGATHGVYAGSCQGPDPGDLSSPLVRGHKEPDQWQYDEKGN 360
QY 403 TTYARGGLALNEFRSIPQ-----VDLIANARKGVR----- 433
Db 361 RTVAEBCRNAKIKTEPKLDAEGKQVYVDEYSGSRTPVYVVDYELDEKGNKIQETNPD 420
QY 434 -----GYSHT-----VATENLDEFGIYK-----STFHPADGLSLIGG 466
Db 421 GTPAFTGSGTVPVWKTWKVADDDHVPALYNYAKYLNNTKTHSLTAGTRFNTGRHLHLOG 480
QY 467 GRLGHYKIESGEGKTL-----HKASKT---KFTGYAGAVVDLN 501
Db 481 --LHYTRYETSQTKDMPVRYQPPASDFQTASSIKADQDHYTAKMQGHKLTGYAGITTYDLT 538
QY 502 DNSSLYLSLSQLYTPQTNLDADGK-LLKPRQGNQFVEYKGSYMDRLNARVSFYRMKDK 560
Db 539 PQQSIYGSYTKIFKQODNDVDSAKTVLPPLVGTNYEYVWKGAFLOGLRLNASFALFYLEQK 598
QY 561 N-----AAAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAQYSYLSHQI 611
Db 599 NRTVVDVGYVPGAGKQSGPQTVAKPIGKVVSRAEFLSGELNEDWKVPAGYTYNKSRY 658
QY 612 KTAS-----NSRDD--GIFLLMPKHSANLWTTYYOV--TPELTIGGVNAMSGITSSA 659
Db 659 KNAAEVNAERLAKNTGADPNFSNFTPVHIFRFGTSPHIPNTGLTVGGVSAQSGTSSLY 718
QY 660 GMHAGGYATPDMAAYRFTPKLQINADNIFNRHY-----YARVGGANTFNIPOSERTWT 715
Db 719 NIROGGYGLIDGFVRYELGKHAKLSLIGTLNLRGTYFENNRYNTRGANNFY--GEPRTVS 776
QY 716 ANLRYSF 722
Db 777 MKLDWQF 783

RESULT 13
AAB68923
ID AAB68923 standard; Protein; 702 AA.
XX
AC AAB68923;
XX
DT 18-APR-2001 (first entry)
XX
DE Neisseria meningitidis protein #22.
XX
KW Meningococcus; meningitis; bacteraemia; vaccine; dsbA; fhaB; fhuA;
KW rni5; rth; tolC.
XX
OS Neisseria meningitidis.
XX
PN EPI069133-A1.
XX
PD 17-JAN-2001.
XX
PF 13-JUL-1999; 99EP-0401764.
XX
PR 13-JUL-1999; 99EP-0401764.
XX
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Nassif X, Tinsley C;
XX
DR WPI; 2001-082916/10.

DR N-PSDB; AAF56463.
 XX Immunogenic polypeptides derived from *Neisseria meningitidis* and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against *Neisseria* infections e.g. bacteraemia and meningitis -
 XX
 XX Claim 3, Fig 22B, 240pp; English.
 XX
 XX The present invention provides the protein and coding sequences of
 CC several genes from *Neisseria meningitidis*. These include the *dsbA*, *fhaB*,
 CC *fhuA*, *rnh15*, *rnh17*, *rnh18*, *rnh19*, *rnh20*, *rnh21* and *tolC* genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX
 XX Sequence 702 AA;
 SQ

Query Match 9.9%; Score 375.5; DB 22; Length 702;
 Best Local Similarity 22.2%; Pred. No. 7.3e-25;
 Matches 171; Conservative 136; Mismatches 336; Indels 127; Gaps 33;

8 FRIMTAATVLAALSSVFAAQTAD-----LEFVHKGQSSVNAIVTEKNGDYSSFA 59
 5 FHIALPLTIIA--SEPVAAADTODNGEHYATLPVSVVGGSDTSVLKGYIN--YDEAA 60
 60 VIVGTRIPASLREIPQSVSIITNQVKNRVDTPDQARPTGLRYLSNDGSSVYARG 119
 61 VTRNGQL---IKETPQITDITLNIQKKNVGTNDLSLLEGNAIGDAAYDMRG--ESIFLRG 116
 120 YETSEYNI--DGLPAQMSINGTLPLNFAPDRVEMRGPSGLFDSGEMGIVNLVYRKP 177
 117 PQADASDIYRDGVBESGQ--VRSTANI--ERVEILKPSVLYGRNGGIVNMSKYA 172
 178 TKAFQHAAGFGTHQYKAEADVSGSLNSDGSVR-----VMAQTVGASPR 225
 173 NFKQSNIGAVYSGWANSRLMNDINEVLNKNVAIRLTGEVGRANSPRSGLDSKKNVVSFS 232
 226 PAEKNNRHETFFVAAADWDINDPTVLGAGLYQQRHLAPYNGLPADANNKLPSPQHVFG 285
 233 ITVKLND-----GLKWT-----GQYTYDVERPDRSPKTSVYDR--GLPYMGF-- 276
 286 ADMNKKRNSHDVF--ADLKHYFGNGGYGKVMRYSDRDADSNVAFAGSKLGMKTPAGRPG 344
 277 AHNNDVVKKQLQVWRSDILEVAFNDKWRACWQLAHRTAOPDFHFGYASENGNLIKENYAW 336
 345 CNTADKACAVGTEIKOKALAFDASYSRPFILGNTANEFVIGAYNRRRSINEGRTT 404
 337 QQT-----DNKTLSSNLTLDGTYTIGREPENHLTVGMDSR-----EHRNPT 377
 405 LVARGGIALNEFRSIPQVDLIANARKGVRSHTVATEN---LDERFGYKSTFHPADGL 461
 378 LGRVGSF-----TVP--INPYDRAWMPASGRLOPILQNHKADSYGIFQNLFSATPDL 430
 462 SLIGGRGLHYKTESGEGKTLHKAASKTKETGVA-----GAYVDLNDNSLYLSQLYTP 516
 431 KFLVGGRYKTYTFNSENKLT---GNSRQYSGHFSFPIGIVNMINPHVTLTAYSYNGCFAP 487
 517 -----GTNLDADOKL---KPRGQGFVGVKGSYMDDLNANVSFPMKDKAAALNRP 568
 488 YGGRGLSIDTSSAAVFNAAPEYTRQETGVKSSWDDRLSTLSAQLERFIRYRDP 547
 569 NNKTRVYALGKRWMEGVETELISGAVTPKQIHAQSYLSLSQIKTASNSRD---GIFLL 625
 548 KNNPIYVAVSGKRRSGVELSAIGQIIPK--KLYLRGSLGMQAKVYEDKENPRVGIHLN 606
 626 MPKH--SANTMTTYQVTPELTIGGVNAMSGITSSAGMHAG-----GYATFDAM 672
 607 NTSNVTGNLFRRY--TPTENLYGEI-----GVGTG--GKRYGNSRNKEVTLTLPFARVDM 659
 673 AAVRFPPKLTQINADNIFRRHYAVGVGANTNINIGSEPTWANKLRYSF 722
 660 LGNNH--KNVAVVTPAANLLNOKYV-----RSDSPMGNPGRGYARVAYRFR 702

RESULT 14
 AAB68921
 ID AAB68921 standard; Protein; 702 AA.
 XX
 XX AAB68921;
 AC
 XX
 XX 18-APR-2001 (first entry)
 DT
 XX
 XX *Neisseria meningitidis* protein #20.
 DE
 XX
 XX Meningococcus; meningitis; bacteraemia; vaccine; *dsbA*; *fhaB*; *fhuA*;
 KW *rnh15*; *rnh*; *tolC*.
 XX
 XX *Neisseria meningitidis*.
 OS
 XX
 XX EPI069133-A1.
 PN
 XX
 XX 17-JAN-2001.
 PD
 XX
 XX 13-JUL-1999; 99EP-0401764.
 PF
 XX
 XX 13-JUL-1999; 99EP-0401764.
 PR
 XX
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA
 XX
 XX (PLAC) MAX PLANCK GBS FOERDERUNG WISSENSCHAFTEN.
 XX
 XX Nasself X, Tinsley C;
 PI
 XX
 XX WPI; 2001-082916/10.
 DR
 XX
 XX N-PSDB; AAF56461.
 DR

Immunogenic polypeptides derived from *Neisseria meningitidis* and the
 PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against *Neisseria* infections e.g. bacteraemia and meningitis -
 XX
 XX Claim 3, Fig 20B, 240pp; English.
 XX

The present invention provides the protein and coding sequences of
 CC several genes from *Neisseria meningitidis*. These include the *dsbA*, *fhaB*,
 CC *fhuA*, *rnh15*, *rnh17*, *rnh18*, *rnh19*, *rnh20*, *rnh21* and *tolC* genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.
 XX

Query Match 9.9%; Score 374.5; DB 22; Length 702;
 Best Local Similarity 22.5%; Pred. No. 9e-25;
 Matches 175; Conservative 134; Mismatches 326; Indels 143; Gaps 35;

8 FRIMTAATVLAALSSVFAAQTAD-----LEFVHKGQSSVNAIVTEKNGDYSSFA 59
 5 FHIALPLTIIA--SEPVAAADTODNGEHYATLPVSVVGGSDTSVLKGYIN--YDEAA 60
 60 VIVGTRIPASLREIPQSVSIITNQVKNRVDTPDQARPTGLRYLSNDGSSVYARG 117
 61 VTRNGQL---IKETPQITDITLNIQKKNVGTNDLSLLEGNAIGDAAYDMRG--ESIFL 113
 118 RGEYSEYNI--DGLPAQMSINGTLPLNFAPDRVEMRGPSGLFDSGEMGIVNLVYRKP 175
 114 RGFQADASDIYRDGVBESGQ--VRSTANI--ERVEILKPSVLYGRNGGIVNMSKYA 169
 176 RPTKAFQHAAGFGTHQYKAEADVSGSLNSDGSVR-----GR-----VMAQTVGAS 223
 170 YANFKQSRNIGAVYSGWANSRLMNDINEVLNKNVAIRLTGEVGRANSPRSGLDSKKNVVSFS 229
 224 PRPAEGNNRHETFFVAAADWDINDPTVLGAGLYQQRHLAPYNGLPADANNKLPSPQHVFG 283
 230 PSTITVLDN-----GLKWT-----GQYTYDVERPDRSPKTSVYDR--GLPYMGF 274
 284 VGADWNRKFNQNSHDVF--ADLKHYFGNGGYGKVMRYSDRDADSNVAFAGSKLGMKTPAGR 342

Db 275 F-AHRNDFVKDKLQVRSDELFAPNDKWAQWQLAHRTAAQDFHPYAGSENGNLIKRYN 333
 Qy 343 PGCNTADDKACAVGLTEIKQKALAFDASYSRPFRLGNTANFVIGADYNRRFRSTNEQGR 402
 Db 334 AQOQT-----DNKTLSSNFTLNGDYTGIREPHNLTVGMDYSR-----EHRN 374
 Qy 403 TTYLARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDEFGIYKSTFHPAD 459
 Db 375 PTLGYRGSP-----TVP-INFPYDRASWPASGRLOPILTONRHKADSYGIFVQNFISATP 427
 Qy 460 GLSLGGGRGLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNDNSLYLSLSOLY 514
 Db 428 DLKFLVGGRYDKYTFENSEKLT---GNSQYSGHGSFSPNIGAVWNPVHTLYASYNKGF 484
 Qy 515 TP-----QTNLDADGKLLKPRQGNQFEVGYKSGYMDRLNARVSYFYMRKDK 560
 Db 485 APYGGRGYLSIDTLSSAVFNAD----PEYTRQYETGVKSSWLDRLSTLSAQAQIERP 539
 Qy 561 NAAAPLNPNKKTRYAALAKRWVMEGVETISGAVTPKQIHAQYSLHSQIKTASNSRDD 620
 Db 540 NTRYRDPKNNPIYAVSGKHSRGVLSAIGQIIPK-KLYLRGSLGVNQAKVVEDKENP 598
 Qy 621 ---GIFLLMPKH-SANLWTTYQVTPELTIGGGYNAMSGITSSAGMHAG-----664
 Db 599 DRVGHLNNTSVNGLFFRY--TPTENLYGEI---GVGTG-GRKYGNSRNKEVTTLP 651
 Qy 665 GYATDAMAARYFTPKLQIADNADIFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
 Db 652 GPARVDAMLGWNH-KNVNVTFAANLFNQKYW-----RSDSMGPNRGYTRAVNRYF 702

RESULT 15
 ABB52462
 ID ABB52462 standard; Protein; 663 AA.
 AC ABB52462;
 DT 11-FEB-2002 (first entry)
 DE Escherichia coli polypeptide SEQ ID NO 263.
 KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
 KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
 KW systemic infection; non-diarrhoeal infection; septicemia;
 KW pyelonephritis; antibiotic resistance.
 OS Escherichia coli.
 XX WO2000166572-A2.
 PN 13-SEP-2001.
 PD 12-MAR-2001; 2001WO-EP03445.
 PF 10-MAR-2000; 2000FR-0003145.
 PR 02-FEB-2001; 2001FR-0001449.
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 PI WPI; 2001-550253/61.
 XX A library of DNA fragments of Escherichia coli strains for the
 PT phylogenic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -
 XX Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
 CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
 CC B2/D+A-. The polynucleotides have potential antiinflammatory.

CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.

XX SQ Sequence 663 AA;

Query Match 9.8%; Score 371; DB 22; Length 663;
 Best Local Similarity 22.1%; Pred. No. 1.7e-24;
 Matches 171; Conservative 121; Mismatches 285; Indels 198; Gaps 32;

Qy 20 ALSSSVFAA-----QTADLETVHKQGSYNAIVTEKNGDYSSFAVTVGKIPASL 70
 Db 15 ALFELLFAAPMIHATDSVTTKDGETIVTADAN---TATEATDGYQLSTSTATLTDMPM 71
 Qy 71 REIPQSVSIITNQOVKDRNVDTPQLARKTPGLRVLNSDDG--RSSVYARGYSEYNID 128
 Db 72 LDIPQVNTVSQVLEQNQATTLDEALYNSNV-VQNTILGGTQDAFVRGFG---GANRD 127
 Qy 129 GLPQMQSINGTLNLP--AFDRVEMRGSGCLFDSGEMGGIVNLVRKTRTKAFOGHAA 186
 Db 128 G-SIMTNGRLTVLPSRNAATERVEVLKGPASTLYGLIDPGGLINVVTKRPEKTFHGSVS 186
 Qy 187 A-----GFGTHKQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEK---NNRHETFY 237
 Db 187 ATSSSFGGTT-----GQLDITGPIEGTQLAYRLTGEVQDEDIWRNFGKERSTFI 235
 Qy 238 A-AADWDINPDVILGAGLYLQORHLAPYNGLPADANNKLPSPQHFVFGADWNKFKMNSH 296
 Db 236 APSLTWFGDNATVTMAEV-----HL-----NSQWT-----260
 Qy 297 DVFADLKHYPGNGGYGKVMRYSDRSDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVG 356
 Db 261 -----ARFDYSYSDKY-----SDNQA-----277
 Qy 357 LGTEIKOKALAFDASYSRPPF-----LGNTANEFVIGADYNRRFRSTNEQGRITLYA--- 407
 Db 278 -----RVTAIDATGTLTRVDATOGSTQRMHATRADL-----QGNVDIAGPYNE 322
 Qy 408 -RGLALNEFRSIPQVDLIA-----NARKGVRGYSHTVATENLDE-----FCIY 450
 Db 323 LLGGVSY-EYVDLLRTDMIRCKKAKDFNIYNPVYNTSKCTTVSASDSQTIKQENYSAY 381
 Qy 451 GKSTFHPADGLSLGGRLGHYKIESGEGKTLHKASKTK---FTGYAGAVYDLNNDNSLY 507
 Db 382 AQDALYLTDMNIAVAGIRYQYTYQYAGKGRPFNVNTDSRDEQMTPKLGLVYKLTTPSVSLF 441
 Qy 508 LSLSOLYTPQNLDAADGKLLKPRQGNQFEVGYKSGYMDRLNARVSYFYMRKDKNAAAPLN 567
 Db 442 ANYSQTFMPOSSIASYIGDLPPSSNAYEYGAKEFELFDG-ITADIALDIHKRNLVLTES 500
 Qy 568 PNNKTRYAALAKRWVMEGVETISGAVTPKQIHAQYSLHSQI-----KTASNSRD 619
 Db 501 IGDETIAKTA-GRVRSRGVEVDLAGALTENINIIASVGYTDAKVLDEPDYAGKFLPN--- 556
 Qy 620 DGIFLLMPKHSAANLWTTYQV-----TPELTIGGGYNAMS---GITSSAGMHAGGYATFDAM 672
 Db 557 -----VPRHTGSLFTYDIHNMFGNNLTLPFGGGGHGVSRRSATNGADYLYLPGYFVADAF 610
 Qy 673 AAYRFTPKLK---LQINADNIPNRHY-ARVGGANTFNIPGSERTWTANLRYSP 722
 Db 611 AAYKM--KLQYPTVLQNLVKNLFDKTYTTSIATNNLGNQIGDPRVQFTVKMEF 663

RESULT 16
 AAB68926
 ID AAB68926 standard; Protein; 703 AA.
 XX


```

AC AAB68926;
XX
XX 18-APR-2001 (first entry)
XX
XX Neisseria meningitidis protein #25.
DE
XX Neisseria meningitidis protein #25.
XX Meningococcus; meningitis; bacteraemia; vaccine; dba, fnaB, fnaH,
KM rna1, rth, toIC.
XX
OS Neisseria meningitidis.
XX
XX EPI069133-A1.
XX
XX 17-JAN-2001.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX 13-JUL-1999; 99EP-0401764.
XX
XX (INM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX Nasif X, Tinsley C;
XX
XX WPI; 2001-082916/10.
XX
XX N-PSDB; AAF56466.
XX
XX
XX Immunogenic polypeptides derived from Neisseria meningitidis and the
XX nucleic acids that encode them, useful for diagnosing and vaccinating
XX against Neisseria infections e.g. bacteraemia and meningitis -
XX
XX Claim 3; Fig 25B; 240P; English.
XX
XX
XX The present invention provides the protein and coding sequences of
XX several genes from Neisseria meningitidis. These include the dba, fnaB,
XX fnaH, rna1, rth17, rth18, rth19, rth20, rth21 and toIC genes. These can
XX be used in the diagnosis and treatment of infection by the bacterium,
XX which can lead to meningitis and bacteraemia, and in vaccines to prevent
XX such infection.
XX
XX
XX Sequence 703 AA;
SQ
Query Match 9.8%; Score 371; DB 22; Length 703;
Best Local Similarity 22.4%; Pred. No. 1.9e-24;
Matches 174; Conservative 131; Mismatches 332; Indels 140; Gaps 33;
QY 8 PRINMATAIVLALSSVPAQAQAD-----LETVHKGQSSVNAIYTERKGDYSSRA 59
DB 5 PHLALPLTLIA--SEPVAAADTDQNGEHYATLPTVSVVGSDTSVLKGYIN--YDEAA 60
QY 60 VTWGTKIPASLREIPQSVSITITNOQVKNVDTFDQARKEPGLRVLSNDDGRSSVYARG 119
DB 61 VTRNGD---IKETPOTITLNIQKKNKVTNLSLLEGNAGIDAIYWRG-ESTIFLKG 116
QY 120 VEYSEYNI--DGLPAQMSINGTLFNLFAFDRAVEVRGSGGLFDSGEMGIVNLVRKEP 177
DB 117 FOADADIVYDVGVRSGQ--VRSTANI---ERYEILKGPSSVLVGRITGGGVINMYSKA 172
QY 178 TYAFQGHAAAGBETHKOYRAEADVSGSLNSDSVR-----VMAQTVGASPR 225
DB 173 NFKQSHNIAVYGSWMANRSLMDINEVLKNVAIRLTGEVGRANSPRSGDSKNVAVSPS 232
QY 226 PAKKNRHETFFVAAADWDINPDTVLGAGYL YQGRHLAPYNGLPADANNKLPSPQHVFPV 285
DB 233 IYVKLDN-----GLKKT-----GQITYDVERTPDRSPFKSYDR-GLPYRMGF- 276
QY 286 ADWNKFKMNSHDVF-ADLKHYFGNGGYGKVMKYSDDRADSNYAFAQSKLGMKTPARGP 344
DB 277 AHNNDVFKDKLQVWRSDLEAYAFNDKWRAGWQLAHRTAQDFHFVYAGSEGNLIKENVAM 336
QY 345 CMTAADKACAVGLGTETIKQKALAFDASYSRPFILGNTAFVGVGADYNNRSTRNBQRTT 404
DB 337 QQT-----DNKTLSSNFTLNGDYITIGRFENHLTVGMDYSR-----EHRNPT 377

```

```

QY 405 L-YARGGLALNEFRSIPVDLLIANKRGVGYSHTVATEN---LDEFGYKSTPHPADG 460
DB 378 LGSRAFTA-----SIDPYDRASWPASGRLOPILTONRHKADSYGIFQNIFFATPD 429
QY 461 LSLIGGRLGHYKIESBEKTLHKASKTFGTGA-----GAYVDLNDNNSTYLSLSQLYT 515
DB 430 LKFVLGGRYDKTYTFNSEKLT--GSSROYSGHSFSPNIGAWVINPVTLLVASYAKGPA 486
QY 516 P-----QTNLDADGKLLKPRQNOFEVGYGSYMDBRLNARVSFPRMKDN 561
DB 487 FYGGRGYISTSSSAVNAD-----PEYTRYETGVKSSWDDRLSTLLSAIQIERN 541
QY 562 AAAPLNPNKKTRYAALGRVMEGVETEISGAVTPRKOIHAQSYLHSGIKTASNSRDD- 620
DB 542 IRYRPEQNDPYTWAAGGGRSGVELSALGQIIPR-KLYLNGSLGVQAKVVEDKEND 600
QY 621 --GIFLMPKH-SANLMTTYQVTPELTIGGVNANSGITSSAGMAG-----G 665
DB 601 RVGIHLNNTSNVTGNLFFRY--TPTENLYGEI---GVGTG-GRKYGNSRNKEVTTLP 653
QY 666 YATFDMAAYRFTPKLQIINADNIFNRHYARVVGANTFNI PGSEBRTWAMLRYSF 722
DB 654 FARVDMLGWNH-KVNVITFAALNLNQKW-----RSDAMPGARITYTARVNTSF 703

RESULT 17
AAV75568
ID AAV75568 standard; Protein; 703 AA.
XX
XX AAV75568;
XX
XX 21-MAR-2000 (first entry)
XX
XX
XX Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2610.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
XX antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
XX antibacterial; gene therapy.
XX
XX
XX Neisseria meningitidis.
XX
XX WO9957280-A2.
XX
XX 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
XX
XX 01-MAY-1998; 98US-0083758.
XX
XX 31-JUL-1998; 98US-0094869.
XX
XX 02-SEP-1998; 98US-0098994.
XX
XX 02-SEP-1998; 98US-0098062.
XX
XX 09-OCT-1998; 98US-0103749.
XX
XX 09-OCT-1998; 98US-0103794.
XX
XX 09-OCT-1998; 98US-0103796.
XX
XX 25-FEB-1999; 99US-0121528.
XX
XX (CHTR ) CHIRON CORP.
XX
XX (GENO-) INST GENOMIC RES.
XX
XX Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX Tettelin H, Veneri JC;
XX
XX WPI; 2000-062150/05.
XX
XX N-PSDB; AA254330.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
XX vaccines and diagnostics -
XX
XX Claim 2; Page 1238; 1453P; English.
XX
XX AA253015 to AA254536, AA254577 to AA254615, and AAV74253 to AAV75941
XX
XX
XX

```


represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and polypeptides. AA25437 to AA25457 and AA25461 to AA25473 represent PCR primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to *Neisserial* bacteria (e.g. meningitis and septicaemia), to detect the presence of *Neisseria* bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

Sequence 703 AA:

Query Match	9.8%	Score 370;	DB 21;	Length 703;
Best Local Similarity	22.4%;	Pred. No. 2.3e-24;		
Matches 174:	Conservative 130;	Mismatches 334;	Indels 138;	Gaps 33;

Qy	8	F R I N M T A A T V L A A L S S V F A A Q T A D - - - - - L E T V H I K G R S Y N A I V T E K N G D Y S S P A	59
Db	5	F H L A L L P T L I I A - - S P F V A A A A T Q D N G E H Y T A T L P T V S V V G Q S D T S V L X G Y I N - - Y D E A A	60
Qy	60	V T V G T K P A S L R E I P O S V S I I T W Q O V K D R N V D F D Q L A R K T P C L R V L S N D D G R S S V A R G	119
Db	61	V T R N G Q L - - - I K E T P Q I D T I N T Q K N K Y C T N D L S S I L E G N A G I D A A Y D M R G - E S I F L R G	116
Qy	120	Y E Y S E Y N I - - D G L P A O M Q S I N G T L P N L F A P D R V E V M R G S G L F D S S G E M G G I V N L V R K P R	177
Db	117	F O A D A S I D Y R G V R E S Q - V R R S T A N I - - E R V E I L K G P S S V L Y G T R N G G V I N V M S K Y A	172
Qy	178	T K A P Q G H A A A G F P H K O Y K A B A D V S G S L N S D G S V R - - - - - G R - - - - - V M A Q T V G A S P R	225
Db	173	N F K Q S R N I G T V Y G S W A N R S L N M D I N E V L N K N V A I R L T G E V G R A N S P R S G I D S K N V M S P S	232
Qy	226	P A E K N N R H E T F Y A A A D W I N P D T V I L G A G Y I Q O R H L A P Y N G L P A D A N N K L P S L P Q H V F V G	285
Db	233	I T V K L D N - - - - - G L K W T - - - - - G Q Y T Y D N V E R T P O R S P T K S V Y D R P - G L P Y R M G F -	276
Qy	286	A D M N K F K W N S H D V F - A D L K H V F G N G G Y K G V M R Y S D R D A D S N T A F A G S K L G M K T P A G R P G	344
Db	277	A H R N D F V K D K L Q W R S D L E Y A F N D K W P A Q W L A H R T A A Q D F D H F Y A G S E N G L N I K E N Y A W	336
Qy	345	C N T A D D K A C A V G L T E T I K Q A L A P D A S Y S R P F R L G N T A N E F V I G A Y N R F R S T N E O G R T T	404
Db	337	Q O T - - - - - D N K T L S N L T L N G D Y T I G R F E N H L T V G M D Y S R - - - - - E H R N P T	377
Qy	405	L Y A R G G L A N E F R S I P O V D L I A N A R K G V R C Y S H T V A T E N - - - L D E F G I Y G K S T F F P A D G L	461
Db	378	L - - - - G F S S A F S A S I N P Y D - - - R A S W P A S G R L O P I L T Q N R H K A D S Y G I F V Q I F A S T P D L	430
Qy	462	S L I G G R L G H Y K T E S G E K T L H K A S K T F T G Y A - - - - - G A V Y D L N D N N S I L Y L S L S O L Y T P	516
Db	431	K F V L G G R Y D K Y T N S E N K L T - - - G S S R Q Y S G H S F S P N I G A V M N I N P V H T L Y A S Y N K G F A P	487
Qy	517	- - - - - Q T N L D A D G K L L K P R Q N Q F E V G Y K G S Y M D D R L N A R V S F Y R M K D K N A	562
Db	488	Y G R G G Y L S I D T L S S A V F N A D - - - - - P E Y T R Q Y E T G V K S W L D D R L S T T S L S A Y Q I E R F N I	542
Qy	563	A A P L N P N K K T R Y A A L K R V M E G V E T I S A V T P K W I H A G Y S L H S O J K T A S N S R D D - - -	620
Db	543	R Y R P D P K N P Y I V A V S G K H R S R G V E L S A I G O I I P K - K L Y L R G S L G V M Q A K V V E D K E N P D R	601
Qy	621	- G I P L L M P K H - S A N L W T Y Q V T P E L T I G G V G N A M S G I T T S A G M H A G - - - - - G Y	666
Db	602	V G I H L N T S V N G L F R Y - - T T E N L Y G E I - - - - - G W T C T - G K R Y C Y D S R N K E V T L L P G F	654
Qy	667	A T F D M A A Y R F T P K L K L O I A D N I F N R H Y Y A R V G G A N T F N I P G S E R T W T A N L U R Y S F	722
Db	655	A R V D M A L G M N H - K N V N V T F A A A L F N O K Y W - - - - - R S D S M P G N P R G Y T A R V N Y R F	703

RESULT 18

AA868916	
ID	AA868916 standard; Protein; 703 AA.
XX	
XX	AA868916;
XX	
DT	18-APR-2001 (first entry)
XX	
XX	Neisseria meningitidis protein #15.
DE	
XX	
KW	Meningococcus; meningitis; bacteraemia; vaccine; ddaA; fhaB; fhuA;
KW	rni5; rth; tolC.
XX	
OS	Neisseria meningitidis.
XX	
XX	EP1069133-A1.
PN	
XX	
PD	17-JAN-2001.
XX	
XX	
PF	13-JUL-1999; 99EP-0401764.
XX	
XX	
PR	13-JUL-1999; 99EP-0401764.
XX	
XX	
XX	
XX	(INRM) INSERM INST NAT SANTE & RECH MEDICALE.
PA	(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX	
XX	Nassif X, Tinsley C;
PI	
XX	
XX	WPI; 2001-082916/10.
DR	N-PSDB; AAP56456.
XX	
XX	
XX	Immunogenic polypeptides derived from Neisseria meningitidis and the
PT	nucleic acids that encode them, useful for diagnosing and vaccinating
PT	against Neisseria infections e.g. bacteraemia and meningitis -
XX	
XX	
PS	Claim 3; Fig 15B; 240pp; English.

AA The present invention provides the protein and coding sequences of
CC several genes from *Neisseria meningitidis*. These include the *dsbA*, *flaB*,
CC *flaA*, *rfi5*, *rth17*, *rth18*, *rth20*, *rth21* and *tolC* genes. These can
CC be used in the diagnosis and treatment of infection by the bacterium,
CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
CC such infection.

Sequence 703 AA:

Query Match	9.8%	Score 370;	DB 22;
Best Local Similarity	22.4%	Pred. No. 2.3e-24;	Length 703;
Matches 174;	Conservative 130;	Mismatches 334;	Indels 138;
Gaps 33;			

Qy	8	FRINMTAATVLAALSSSVFAAQATAD-----LETVHIKGQBSYNAIVTEKNGDYSSFA	59
Db	5	PHLALLPTLIIA--SPFVAAADTODNGEHVYATLPTVSVVGQSDTSVLKGYIN--YDEAA	60
Qy	60	VTVGTKIPASLREIPQSVSIITNQOVKDRNVDTFDQLAKTPGLRVLSDNDGRSSVYARG	119
Db	61	VTRNGQL---IKETPOTIDTLNQKKNYGTNDLSSILEGNAGIDAAYDMRG-ESIFLGR	116
Qy	120	YEYSEYNI--DGLPAQWOSINGTLPNLPAFDRVEMRGPSGLPDSGSEWGIYNLVRKRP	177
Db	117	FQADASDIYRDGVRESQG-VRRSTANI---ERVEILKGPSSVLVYGRNTGGGVINMWYSKYA	172
Qy	178	TKAPQGHAAAGFGTHKQYKAEADVSGSLNSDGSVR-----VMACTVCGASPR	225
Db	173	NFKOSRNIGTVYGSWANRSLNMDINEVLNKNQVAIRLTGEVGRANSFRSGIDSKNVWVSPS	232
Qy	226	PAEKNRHHETFYAAADWDINPDTVLGAGYLYOORHLAPYNGLPADANNKLPSLPQHVFVG	285
Db	233	ITVKLDN-----GLKWT-----GQYTDNVERTPDRSPKTSVYDRF-GLPYRMGF-	276
Qy	286	ADWNKFKQNRHSHVFP-ADLKHYFGNGGYKVKVMYRSRDDADSNVAFAGSKLGMKMTPAGRPG	344
Db	277	AHRNDFYKDLQVWRSDLEYAFNDKWRAOQWLAAHRTAAQDFDHIFYAGSENGNLIKRNYAW	336

QY 345 CNTADKACAVGLGTEIKOKALAFDASYSRPRFLGNTANEFYLGATYNNFRSTNECGRTT 404
 Db 337 QQT-----DKRTLSSNLTLLNGDYTIGRFENHLLTVGM DYSR-----EHRNPT 377
 QY 405 LYARGGLALNEFRSIPQVLLINARKGVAGSHYATVEN---LDERGIYKSTFHPADGL 461
 Db 378 L----GFSSAFSASINPYD---RASWPASGRLOPILTONRHRADSIGIFQNI FSA TPDL 430
 QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLSLYLSOLYTP 516
 Db 431 KRYLGRYKTYFNSENKLT---GSSRQYSGHSFSPNIGAVNNINVHLLYASYNNGFAP 487
 QY 517 -----QTNLDADGKLKPRGQNOPEVGYKSYWDDRLNARVSFRMKDKKA 562
 Db 488 YGGRGGYSLIDTLSSAVFNAD-----PEYTRYETGVKSSWLDRLSTLLSAVQIERFNI 542
 QY 563 AAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAGSYLHSQIKTASNSDD-- 620
 Db 543 RYRPDPKNNPYIYAVSGKRSRGVELSAIGQIIPK-KLYLRGSLGVMAQKVVEDKENPDR 601
 QY 621 -GIFLLMPKH-SANLWTTYQVTPELTIGGVANMAGSITSSAGMHAG-----GY 666
 Db 602 VGHLNNTSNVTGNLFERR--TPTENLYGEI---GVGTG-GKRYGYDSRNKEVTTLPDF 654
 QY 667 ATFDMAAAYRTPPKLKLQINADNIFNRHYIYAVGANTNIPGSERTWTANLRYSF 722
 Db 655 ARVDAMLGNWH-KNVVVTFAAANLIFNOKYW-----RSDSMGPNRGYARVNYRF 703

RESULT 19

AA75567
 ID AA75567 standard; Protein; 703 AA.

AC AA75567;

DT 21-MAR-2000 (first entry)

DE Neisseria meningitidis ORF 761 protein sequence SEQ ID NO:2608.

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.

OS Neisseria meningitidis.

PN MO9957280-A2.

XX 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 98US-0094869.

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.

PR 09-OCT-1998; 98US-0103749.

PR 09-OCT-1998; 98US-0103794.

PR 09-OCT-1998; 98US-0103796.

PR 25-FEB-1999; 99US-0121528.

XX (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
 PI Petersen U, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelein H, Venter JC;

DR WPI; 2000-062150/05.
 DR N-PSDB; AA54329.

XX Novel Neisserial polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics
 XX

PS Claim 2; Page 1237; 1453pp; English.
 XX AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccines, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria meningitidis* (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria meningitidis*, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.

XX Sequence 703 AA.

Query Match 9.7%; Score 368; DB 21; Length 703;

Best Local Similarity 22.4%; Pred. No. 3.5e-24;

Matches 174; Conservative 130; Mismatches 334; Indels 138; Gaps 33;

QY 8 FRINMTAAVTLAASSVAAQTAD-----LEVHMKQGRSYATVTEKQGYSSFA 59
 Db 5 FHIALPLTLIIA--SPVAAADTODGHEHYATLPIVSVVSGSDTSYKGYN--YDEAA 60
 QY 60 VYVGTGIPASLEIPQSVIITNQGVKDNVDPDQARKTGRLVLSNDDGRSSVYARG 119
 Db 61 VTRNGQL---IKETPOTITLINIQKKNYGTNDLSILEGNAGIDAAVYMRG-ESIFLRG 116
 QY 120 YEYSEYNI--DGLPAQMGISNGTLPNLFADREYVWRGSGSFDSSGEMGIYLVKRP 177
 Db 117 FQADSDIYRDGVRESCQ--VRRSTANI---ERVEILKGPSSVLYGRTNGGAYINWVSXYA 172
 QY 178 TKAFQGHAAAGFTKHQYAEADVSGSLNSDGSVR-----VMAQTVASR 225
 Db 173 NFKQSRNIGAVYGSWANRSLNMDINEVINKVAIRLTGEVGRANSFRSGIDSKVAVVSPS 232
 QY 226 PAKKNRHETFFAAADWDINDPVLGAGVLYOORHLPANGLPADANNTLPSPQVFPVG 285
 Db 233 IYVKLDN-----GLKMT-----GQYTYDNVERTPDRSPKSYVDYF-GLPYRMGF- 276
 QY 286 ADWNKFKQNSHDVF-ADLKHVFGNGGYGVKMKYSDRDADSNVAFAPAGSKLGMKTPAGRPG 344
 Db 277 AHRNDFVKDKLVWMSDLEAYAFNDKWRQAQWLHRTAADPDHFPVYAGSNGMLIKXNYAM 336
 QY 345 CNTADKACAVGLGTEIKOKALAFDASYSRPRFLGNTANEFYLGATYNNFRSTNECGRTT 404
 Db 337 QQT-----DKRTLSSNLTLLNGDYTIGRFENHLLTVGM DYSR-----EHRNPT 377

QY 405 LYARGGLALNEFRSIPQVLLINARKGVAGSHYATVEN---LDERGIYKSTFHPADGL 461
 Db 378 L----GFSSAFSASINPYD---RASWPASGRLOPILTONRHRADSIGIFQNI FSA TPDL 430
 QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLSLYLSOLYTP 516
 Db 431 KRYLGRYKTYFNSENKLT---GSSRQYSGHSFSPNIGAVNNINVHLLYASYNNGFAP 487
 QY 517 -----QTNLDADGKLKPRGQNOPEVGYKSYWDDRLNARVSFRMKDKKA 562
 Db 488 YGGRGGYSLIDTLSSAVFNAD-----PEYTRYETGVKSSWLDRLSTLLSAVQIERFNI 542
 QY 563 AAPLNPNKKTRYAALGKRVMEGVETEISGAVTPKQIHAGSYLHSQIKTASNSDD-- 620
 Db 543 RYRPDPKNNPYIYAVSGKRSRGVELSAIGQIIPK-KLYLRGSLGVMAQKVVEDKENPDR 601
 QY 621 -GIFLLMPKH-SANLWTTYQVTPELTIGGVANMAGSITSSAGMHAG-----GY 666
 Db 602 VGHLNNTSNVTGNLFERR--TPTENLYGEI---GVGTG-GKRYGYDSRNKEVTTLPDF 654
 QY 667 ATFDMAAAYRTPPKLKLQINADNIFNRHYIYAVGANTNIPGSERTWTANLRYSF 722
 Db 655 ARVDAMLGNWH-KNVVVTFAAANLIFNOKYW-----RSDSMGPNRGYARVNYRF 703


```
Db 277 AHRNDFVKDLQWRSDELEYAENDKWAQWLAHRTAAQDFHFFYAGSENGNLIKRNAYW 336
Qy 345 CNTADDKACAVGLGTEIKQALAFDASVSRPRLGNANEFVIGADYNNRFRSTNQGRRT 404
Db 337 QQT-----DNKTLSSNFTLNGDYTIQGFENHLTVGMDYSR-----EHRNPT 377
Qy 405 LVARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDFEIGYKSTFFHPADGL 461
Db 378 L-----GFRNFASIDPYD---RASRPASGRQLQRIIAQDRHKADSYGIFVQIFSATPDL 430
Qy 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----CAVYDLNDDNSLYLSLSQLYTP 516
Db 431 KPVLGGRYDKYTFNSENKLT---GSSRQSYGHSFSPNIGAVNNINPVHTLYASYNKAFA 487
Qy 517 -----QTNLDADGKLLKPRQGNQFQEVGYKGSYMDRLNARVSFYRMKDKNNA 562
Db 488 YGGRGGYLSINTSSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLSAYQIERNI 542
Qy 563 AAPLPNNKTRYAALGKRVMEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNRDD-- 620
Db 543 RYRPDQNDPYTWAVGGKHSRSGVELSAIGQIIPK-KLYLRGSLGVQAKVVEDKKNPDR 601
Qy 621 -CIFLLMPKH-SANLWTTTYQVTPELTIGGGVNMAGSITSSAGMHAG-----GY 666
Db 602 VGIHLNNTSNVTGNLFRRY--TPTENLYGEI---GVGTG-KGRYGYNSRNKEVTTLPGF 654
Qy 667 ATFDAMAAVRFPPKLLQIINADNIFNRHYVYARVGGANTFNIPGSERTWTANLRSYF 722
Db 655 ARVDAMLGNWH-KNVNVTFAANLFNQKYW-----RSDSMFGNPRGYTARVNYRF 703
```

RESULT 23

AAB68922
ID AAB68922 standard; Protein; 702 AA.

XX AAB68922;

AC AAB68922;

DT 18-APR-2001 (first entry)

XX Neisseria meningitidis protein #21.

XX Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;

KW rni5; rth; toIC.

XX Neisseria meningitidis.

XX EP1069133-A1.

PN 17-JAN-2001.

XX 13-JUL-1999; 99EP-0401764.

XX 13-JUL-1999; 99EP-0401764.

XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Nassif X, Tinsley C;

XX WPI; 2001-082916/10.

DR N-PSDB; AAF56462.

XX Immunogenic polypeptides derived from *Neisseria meningitidis* and the
PT nucleic acids that encode them, useful for diagnosing and vaccinating
PT against *Neisseria* infections e.g. bacteraemia and meningitis -

XX Claim 3; Fig 21B; 240pp; English.

XX The present invention provides the protein and coding sequences of
CC several genes from *Neisseria meningitidis*. These include the *dabA*, *fhaB*,
CC *fhuA*, *rni5*, *rth17*, *rth18*, *rth19*, *rth20*, *rth21* and *toIC* genes. These can
CC be used in the diagnosis and treatment of infection by the bacterium,

CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
CC such infection.

XX SQ Sequence 702 AA;

Query Match 9.7%; Score 364.5; DB 22; Length 702;

Best Local Similarity 22.8%; Pred. No. 7.3e-24;

Matches 177; Conservative 129; Mismatches 330; Indels 141; Gaps 35;

Qy 8 FRINMTATVLAALSSSVFAAQATD-----LETVHIKQORSYNAIVTEKNGDYSSFA 59

Db 5 FHLALLPLIIA--SFPVAAADTDNGEHYATLPTTVSVVGQSDTSVLKGIN--YDAA 60

Qy 60 VTVGTKIPASUREIPQSVSIITNOQVKNRVNVDTFDQLARKTPGLRVLNSDDGRSSVYARG 119

Db 61 VTRNGQL---IKETPQTIDTLNIQKNKNGYTNLDLSSILEGNAGIDAAAYDMRG-ESIFLRG 116

Qy 120 YEYSEYNI--DGLPAQMQSINGTLPNLPAPDRVEMRCPGLFDSSGEMGGIVNLVRKP 177

Db 117 FOADASDIYRGVRESGQ-VRRSTANI---ERVELKGPSSVLYGRTNGGGVINVMSKYA 172

Qy 178 TKAFQGHAAAGFGTHKOYKAADVSGSLNSDGSVR-----GR-----VMAQTVGASPR 225

Db 173 NFKOSRNIGAVYGWNR-SLNMDINEVLNKNVAIRLTCEVGRANSFRSGIDSKNMVWSPS 231

Qy 226 PAEKNRHETFYAAADWDINDPTVLGAGYLYQQRHLAPYNGLPADANNKLPSPQHFVVG 285

Db 232 ITVKLDN-----GLKWT-----GQTYDNDVERTPDRSPTKSVYDRP-GLPYRMGF- 275

Qy 286 ADMNKFKNSHDVF-ADLKHVFGNGYKVGCMRYSDRDADSNYAFAGSKLGMKTPAGRPG 344

Db 276 AHRNDFVKDKLQWRSDELEYAFNDKWAQWLAHRTAAQDFHFFYAGSENGNLIKRNAYW 335

Qy 345 CNTADDKACAVGLGTEIKQALAFDASVSRPRLGNANEFVIGADYNNRFRSTNQGRRT 404

Db 336 QQT-----DNKTLSSNFTLNGDYTIQGFENHLTVGMDYSR-----EHRNPT 376

Qy 405 L-YARGGLALNEFRSIPQVDLIANARKGVGYSHTVATEN---LDFEIGYKSTFFHPADG 460

Db 377 LGYNRAFSA-----SINPYD---RASWPASGRQLQPILTQNRHKADSYGIFVQIFSATPD 428

Qy 461 LSLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNDDNSLYLSLSQLYT 515

Db 429 LKFLVGGRYDKYTFNSENKLT---GSSRQSYGHSFSPNIGAVNNINPVHTLYASYNKGFA 485

Qy 516 P-----QTNLDADGKLLKPRQGNQFQEVGYKGSYMDRLNARVSFYRMKDKN 561

Db 486 PYGGRGGYLSIDTSSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLSAYQIERFN 540

Qy 562 AAPLPNNKTRYAALGKRVMEGVETEISGAVTPKWQIHAGYSYLHSQIKTASNRDD- 620

Db 541 IRYRPDQNDPYTWAVGGKHSRSGVELSAIGQIIPK-KLYLRGSLGVQAKVVEDKKNP 599

Qy 621 --CIFLLMPKH-SANLWTTTYQVTPELTIGGGVNMAGSITSSAGMHAG-----G 665

Db 600 RVGIHLNNTSNVTGNLFRRY--TPTENLYGEI---GVGTG-KGRYGYDSRNKEVTTLPG 652

Qy 666 YATFDAMAAVRFPPKLLQIINADNIFNRHYVYARVGGANTFNIPGSERTWTANLRSYF 722

Db 653 FARVDAMLGNWH-KNVNVTFAANLFNQKYW-----RSDSMFGNPRGYTARVNYRF 702

RESULT 24

AAB68919

ID AAB68919 standard; Protein; 703 AA.

XX AAB68919;

AC AAB68919;

XX 18-APR-2001 (first entry)

DT Neisseria meningitidis protein #18.

DE Meningococcus; meningitis; bacteraemia; vaccine; dsba; fhaB; fhuA;

XX

XX

QY 60 VTGKIPASLREIPQSVIIITNOQVDRNVDFTDQARKTPGLRLVLSNDDGRSSVYARG 119
 DB 61 VTRNGQL---IKETPQITDITLNIQKNKNGTNDLSILEGNAGIDAAVDMRG-ESIFLUR 116
 QY 120 YEYSEYNI--DGLPAQMQSINGTLNLPFAFDEVMRGPGLFDSGEMGGIVNLVRKRP 177
 DB 117 FOADASDIYRDGVRESGQ-VRRSTANI---ERVELKGPSSVLYGTNGGGVINWVKYA 172
 QY 178 TKAFQCHAAAGFGTHKQYKAEADVSGSLNSDGSVR-----VMAQTVCASPR 225
 DB 173 NFKQSRNICTVYGVSWANRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVMWSPS 232
 QY 226 PAEKNNRHETFAAADWDINPDTVLGAGLYQORHLAPYNGLPADANNKLPSPHQHVPVG 285
 DB 233 ITVKLDN-----GLKWT-----GOYTDNVERTPDRSPKSVYDRF-GLPYRMGF- 276
 QY 286 ADMNFKMNSHDVF-ADLKHFGNGGKVGVMRYSDRDADSNYAFAGSKLGMKTPAGRPG 344
 DB 277 AHRNDFVKDLQVWRSDLEYAFNDKWAQQLAHTAAQDFHFTAGSENGNLIKRNVAW 336
 QY 345 CNTADDKACAVGLGTEIKOKALAFDASYRPFRLGNANTANEVIGADYNRFRSTNEQGRRT 404
 DB 337 QQT-----DNKTLSSNLTNGDYTIGRFENHLTVGMDYSR-----EHRNPT 377
 QY 405 LYARGGLALNEFRSIPQVDLIANARKGVRCYSHTVATEN---LDEFGYKSTFHPADGL 461
 DB 378 L-----GFSSAFSASINPYD---RASWPASGRLOPILQNRHKADSYGIFVQNIIFSATPDL 430
 QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLSLSQLYTP 516
 DB 431 KPVLGGRYDKYTFNSENKLT---GSSRQYSGHSFSPNIGAVWNIINPVHTLYASYNKGFP 487
 QY 517 -----QTNLDADGKLLKPRQGNQFVGYKGSYMDRLNARVSPYRMKDKN 562
 DB 488 YGGRGYLSIDTLSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLTSAYQIERFNI 542
 QY 563 AAPLNPNKKTTRYAALGKRVMEGVETEISGAVTPKQIHAGY 604
 DB 543 RYRDPKNNPIYAVSGKHSRGVELSAIGQIPKLTLSARF 584

RESULT 26

AA68925
 ID AA68925 standard; Protein; 697 AA.

XX
 AC AA68925;

XX
 DT 18-APR-2001 (first entry)

XX
 DE Neisseria meningitidis protein #24.

XX
 KW Meningococcus; meningitis; bacteraemia; vaccine; deba; fhaB; fhuA;
 KW rnl5; rth; tolC.

XX
 OS Neisseria meningitidis.

XX
 PN EP1069133-A1.

XX
 PD 17-JAN-2001.

XX
 PF 13-JUL-1999; 99EP-0401764.

XX
 PR 13-JUL-1999; 99EP-0401764.

XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX
 PI Naesif X, Tinsley C;

XX
 DR WPI; 2001-082916/10.

XX
 DR N-PSDB; AAF56465.

XX
 PT Immunogenic polypeptides derived from Neisseria meningitidis and the

PT nucleic acids that encode them, useful for diagnosing and vaccinating
 PT against Neisseria infections e.g. bacteraemia and meningitis -
 XX
 PS Claim 3; Fig 24B; 240pp; English.

XX
 CC The present invention provides the protein and coding sequences of
 CC several genes from Neisseria meningitidis. These include the dbaA, fhaB,
 CC fhuA, rnl5, rth17, rth18, rth19, rth20, rth21 and tolC genes. These can
 CC be used in the diagnosis and treatment of infection by the bacterium,
 CC which can lead to meningitis and bacteraemia, and in vaccines to prevent
 CC such infection.

XX
 SQ Sequence 697 AA;

Query Match 8.4%; Score 316.5; DB 22; Length 697;
 Best Local Similarity 22.1%; Pred. No. 1.7e-19;
 Matches 142; Conservative 107; Mismatches 286; Indels 107; Gaps 24;

QY 8 PRINMTAATVLAALSSSVFAAQTAD-----LETHVIKQQRYSNAIVTEKNGDYSSFA 59

DB 5 FHLALLPTLIIA--SFPVAAADTQDNGEHTATLPTVSVVQSDTSVLKGYIN--YDEAA 60

QY 60 VTGKIPASLREIPQSVIIITNOQVDRNVDFTDQARKTPGLRLVLSNDDGRSSVYARG 119

DB 61 VTRNGQL---IKETPQITDITLNIQKNKNGTNDLSILEGNAGIDAAVDMRG-ESIFLUR 116

QY 120 YEYSEYNI--DGLPAQMQSINGTLNLPFAFDEVMRGPGLFDSGEMGGIVNLVRKRP 177

DB 117 FOADASDIYRDGVRESGQ-VRRSTANI---ERVELKGPSSVLYGTNGGGVINWVKYA 172

QY 178 TKAFQCHAAAGFGTHKQYKAEADVSGSLNSDGSVR-----VMAQTVCASPR 225

DB 173 NFKQSRNICTVYGVSWANRSLNMDINEVLNKNVAIRLTGEVGRANSFRSGIDSKNVMWSPS 232

QY 226 PAEKNNRHETFAAADWDINPDTVLGAGLYQORHLAPYNGLPADANNKLPSPHQHVPVG 285

DB 233 ITVKLDN-----GLKWT-----GOYTDNVERTPDRSPKSVYDRF-GLPYRMGF- 276

QY 286 ADMNFKMNSHDVF-ADLKHFGNGGKVGVMRYSDRDADSNYAFAGSKLGMKTPAGRPG 344

DB 277 AHRNDFVKDLQVWRSDLEYAFNDKWAQQLAHTAAQDFHFTAGSENGNLIKRNVAW 336

QY 345 CNTADDKACAVGLGTEIKOKALAFDASYRPFRLGNANTANEVIGADYNRFRSTNEQGRRT 404

DB 337 QQT-----DNKTLSSNLTNGDYTIGRFENHLTVGMDYSR-----EHRNPT 377

QY 405 LYARGGLALNEFRSIPQVDLIANARKGVRCYSHTVATEN---LDEFGYKSTFHPADGL 461

DB 378 L-----GFSSAFSASINPYD---RASWPASGRLOPILQNRHKADSYGIFVQNIIFSATPDL 430

QY 462 SLIGGRLGHYKIESGEGKTLHKASKTKFTGYA-----GAVYDLNNSLSLSQLYTP 516

DB 431 KPVLGGRYDKYTFNSENKLT---GSSRQYSGHSFSPNIGAVWNIINPVHTLYASYNKGFP 487

QY 517 -----QTNLDADGKLLKPRQGNQFVGYKGSYMDRLNARVSPYRMKDKN 562

DB 488 YGGRGYLSIDTLSSAVFNAD-----PEYTRQYETGVKSSWLDRLSTLTSAYQIERFNI 542

QY 563 AAPLNPNKKTTRYAALGKRVMEGVETEISGAVTPKQIHAGY 604

DB 543 RYRDPKNNPIYAVSGKHSRGVELSAIGQIPKLTLSARF 584

RESULT 27

ABB52899
 ID ABB52899 standard; Protein; 721 AA.

XX
 AC ABB52899;

XX
 DT 11-FEB-2002 (first entry)

XX
 DE Escherichia coli polypeptide SEQ ID NO 1209.


```

KW Escherichia coli; B2/D+A; antiinflammatory; antibiotic;
KM immunosuppressive; extra-intestinal infection; phylogeneity; meningitis;
KM systemic infection; non-diarrhoeal infection; septicæmia;
KM pyelonephritis; antibiotic resistance.
XX
XX Escherichia coli.
OS
XX WO200166572-A2.
PN
XX 13-SEP-2001.
PD
XX 12-MAR-2001; 2001WO-EP03445.
PF
XX 10-MAR-2000; 2000FR-0003145.
PR
XX 02-FEB-2001; 2001FR-0001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA
XX Bingen E, Bonacorsi S, Clermont O, Naasif X, Tinsley C;
PI
XX WPI; 2001-550253/61.
DR
XX
XX A library of DNA fragments of Escherichia coli strains for the
PT phylegenic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A-
XX
XX Example 6; Fig 6; 646pp; English.
PS
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA889533)
CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicæmia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
XX Sequence 721 AA;
SQ
XX
XX Query Match 8.3%; Score 313.5; DB 22; Length 721;
XX Best Local Similarity 22.5%; Pred. No. 3.4e-19;
XX Matches 171; Conservative 127; Mismatches 344; Indels 119; Gaps 33
XX
OY 9 RINMTAAIVLAAISSV---FAOQTADLETYHIKQORSYNAIVTEKNDYSSFAVTVGT 64
DB 2 KTIQIFPALLPALAFIPILPHAAHASTSEDEMI-VTGNTAADTTDSAAAGFGFTINDIDVGP 60
OY 65 KIPADLREIPQSVSIITNOQVADRNVDTFDQARKTPGILRLSN--DGRSSVYARGYE 121
DB 61 LGTKSWIEFPYSTTVYTEMENQOASVSEMLKTSPTOMQARQGMVDGRQO--SRGQO 118
OY 122 YS---EYNIIDG-----PAOMOSINGTLPNLFAFDVEVWRGPGSLFDSGEMGCVI 170
DB 119 GSVVANSRLDGINIVSTYAFVEM-----LEKMDVNLNLTGMLYGPASPAGQGF 166
OY 171 NLVRRKPTKAFQGHAAAGFTHKQYKEADVSGSLNSDGSVGRVVAQTVGASPREAKN 230
DB 167 NFVAKRPFEETLRKVTYLGVQSRSAFTGHADLGHPDENKRFGFRVNLIDQEBEGNVDSBT 226
OY 231 NRHEFFVAAADNDINDPYVL--GAGYLY-QQRHLAPYVGLPADANNKLPISLP----QHY 282
DB 227 LRRKLVSVALLDNINIOFGTQLQDASHYEIIOGVYVGSFNYGP---NVKLPSPAHPKDKML 283
OY 283 FVCGADMNKPKKNSHDVFADLKHYFG-----NGGYGVGKRSYDRDADS-----NYAPA 330
DB 284 ALSTAGN--DLTTDTITSTLYIHFNDDWSMANG--VGGQOADRMRKSYSSKILNNQDGI 338
OY 331 GSKLGAKTTPAG--RPGCNTADDKACAVGSGTEIKOKALAFDASYSRPFRLGNTANEFVIG 388

```

Db 339 SRSMQDSFHAGRFVLSTNTA-----GLNGHIDTGSIGHLSLSTTGCVWMSLYSAGKTG 391
Qy 389 ADYNEFRSTNEQGRTTLTYARGALANEFERSIPQVDLIAMARKGVRSHTVATENLDEFG 448
Db 392 SSSYS-----WGTTNMYHPD--AIIDE-----QQD--GKIRTGGRRSSVNTQOSVTLG 435
Qy 449 IYGKSTFHPADGLSIGGGRLGHYKIBESGEKTIHKASKTFGTGAACAVDYLDNNNSLYL 508
Db 436 --DTVTFPQWSAMEFYLSQSWMLOTRKNYDKHQNTQNQVDENGILSPNALMRYKITPTNAVYV 493
Qy 509 ----SLSQLYTPQT--NLDAADGKLKPRQGNQFEYGVGYGSYMDRLNARVSFFYRMKDGA 562
Db 494 SYADSLEOGGTAIPDESAYKANQOTLPKRSKQYEGLASDI--GMNILGAALFRL--ERP 549
Qy 563 AAPLNPNKKRTRYAALGRVMEGVETTELISCAVTRPMQIHAGYSYLHSQIK-TASNSRDDG 621
Db 550 FAYLDITDV---YKEQGNQVANNGETLAGVNWQGINIYSVTFPLDPKLKOTANASTSNK 606
Qy 622 IFLLMPKHSAWLMTWTYQV--TPELTIGGVN----AMSGITSSAGMHAGGYATFDMAAA 674
Db 607 QVVGVPKVQANMLLAESYSPISPIEWVYSANVAHYTGKRAANDTNTS--YASSYTTWD--LG 661
Qy 675 YRFPPKLR-----LQINADNIENRHYYARVCGATFNPGS 710
Db 662 TRYTKVSNVPTFRVVNVNFDRHYMASIPPSSGTGDGNGS 702

RESULT 28
AAB68918
ID AAB68918 standard; Protein; 700 AA.
XX
XX AAB68918;
XX AC
XX DT 18-APR-2001 (first entry)
XX DE Neisseria meningitidis protein #17.
XX DS
XX KW Meningococcus; meningitis; bacteraemia; vaccine; dsba; fnab; fhuA;
XX rm5; rth; tolC.
XX OS Neisseria meningitidis.
XX PN EPI069133-A1.
XX PD 17-JAN-2001.
XX PF 13-JUL-1999; 99EP-0401764.
XX PR 13-JUL-1999; 99EP-0401764.
XX PA (INRM) INSEEM INST NAT SANTE & RECH MEDICALE
XX PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX PI Nassif X, Tinsley C;
XX DR WPI; 2001-082916/10.
XX N-PADB; AAF56458.

Immunogenic polypeptides derived from Neisseria meningitidis and the nucleic acids that encode them, useful for diagnosing and vaccinating against Neisseria infections e.g. bacteremia and meningitis -
Claim 3; Fig 17B; 240P; English.

The present invention provides the protein and coding sequences of several genes from Neisseria meningitidis. These include the dsba, fnab, fhuA, rm5, rth17, rth18, rth19, rth20, rth21 and tdc genes. These can be used in the diagnosis and treatment of infection by the bacterium, which can lead to meningitis and bacteraemia, and in vaccines to prevent such infection.

Sequence 700 AA;
XX
XX


```
Query Match      8.1%; Score 307.5; DB 22; Length 700;
Best Local Similarity 21.9%; Pred. No. 1.1e-18;
Matches 138; Conservative 109; Mismatches 285; Indels 99; Gaps 24;

QY 8 FRINTAATVLAALSSVFAAQTAD-----LETVHIKQORSYNAIVTEKNGDYSSFA 59
DB 5 FHLALPTLIIA--SPVAAADTDNGEHTATLPTVSVGSDTSVLKGYIN--YDEAA 60
QY 60 VTVGTRKIPASLREIPOSVSIITNQVKDRNVDTPDQARKTFGLRVLNSDDGRSSVYARG 119
DB 61 VTRNGOL---IKETPTIDTLANIQKNKVTGNDLSSILEGNAGIDAAYDMRG-ESIFLRG 116
QY 120 YEYSEYNI--DGLPAQMQSINGTLNLPADPVEVMRGPGLFDSSGEMGGIVNLVRKP 177
DB 117 FOADASDIYRDGVRESGQ--VRRSTANI---ERVEILKGPSSLYLRTGRTGGGVINMVSKYA 172
QY 178 TKAFQGHAAAGCTHKQYKAEADVSGSLNSDGSVR-----GR-----VMAQTVGASPR 225
DB 173 NFKQGRNIGAVYGSRRANRSLNMDINEVLNKNVAIRLTGEVGRANFRSGIDSKNVWVSPS 232
QY 226 PAEKNNRHETFYAAADWDINPDTVLGAGLYQOORHLAPYNGLPADANNKLPSPQHFVFG 285
DB 233 ITVKLDN-----GLKWT-----GQTYDNVERTPDRSPTKSVYORF-GLPYRMGF- 276
QY 286 ADWNPKNNSHDVF-ADLKHVFGNGYGVKGMRYSDRSDADSNVAFAGSKLGMKMTAGRPG 344
DB 277 AHRNDFVKDLQVWRSDELYAFNDKWRAQWLAAHRTAAQDFHFTAGSENGNLIKRNVAY 336
QY 345 CNTADDKACAVGLGTEIKQALAFDASYSPFRLGNTANEFVIGADYRFRSTNEQGRIT 404
DB 337 QQT-----DNKLTSSNFTLNGDYTIQGFENHLLTVGMDYR-----EHRNPT 377
QY 405 L-YARGGLALNEFRSIPQVDLIANARKVGRYSHTVATEN---LDEFGYIGKSTFHPADG 460
DB 378 LGYSRAFTA-----SIDPYDRASWPASGRQLPILTONRHKADSYGLFVQNIISATPD 429
QY 461 LSLIGGRGLGHYKISGEKGLTHKASKTFTGYA-----GAYVDLNDNNSLYLSQLYT 515
DB 430 LKFLVGGRYDKYTFNSENKLT---GSSROYSGHSFSPNIGAVWNINPVHTLYASNGKFA 486
QY 516 P-----QTNLDADGKLL---KPRQNOPEVGYKGYMDRLNARVSVFVKDKKAAAPL 566
DB 487 PYGGGGYLSIDTSSSAVFNADPEYTRQYETGVKSWLNDRLDITLSAYQIBERFNIYRP 546
QY 567 NPNKTKRYAALGKRVMEGVETEISGAVTPK 597
DB 547 DAENNPYTHAVGGKHSRGSVELSAGIIPK 577

RESULT 29
AAR75366
XX AAR75366 standard; Protein; 696 AA.
AC AAR75366;
XX
DT 13-MAY-1996 (first entry)
XX
DE Adhesin.
XX
KW adhesin; plasmid pear; vector; vaccine; intestine colonisation.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT Misc-difference 293
FT /note= "deduced residue from nucleotide sequence
FT is His, this differs from Seq ID5, residue
FT is Gly"
XX
PN W09600233-A1.
XX
PD 04-JAN-1996.
```

```
XX 07-JUN-1995; 95WO-US06994.
XX 24-JUN-1994; 94US-0265714.
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX (UNIW ) UNIV WASHINGTON.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX Besser TE, Bilge SS, Tarr PI, Vary JC;
XX WPI; 1996-068826/07.
XX N-PSDB; AAT10105.
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
XX isolated on plasmid pSC (overlap), for use as a vaccine to prevent
XX bacterial colonisation of bovine intestine
XX Disclosure; Page 31-32; 42pp; English.
XX A candidate adhesin (AAR75366) was identified that is a homologue
XX of the IrgA protein of Vibrio cholerae. The adhesin enables
XX Escherichia coli O157:H7, an antibiotic-resistant, virulent and
XX common food-borne pathogen, to adhere to epithelial cells.
XX Recombinant adhesin was obtd. by expression of the encoding sequence
XX (see AAT10105) in E. coli HB101 (pear). The adhesin can be used as a
XX vaccine for immunisation of cattle against disease or colonisation
XX of mucosal surfaces by O157:H7, thus increasing the safety of food
XX derived from cattle.
XX SQ Sequence 696 AA;
XX
Query Match      7.8%; Score 294.5; DB 17; Length 696;
Best Local Similarity 20.7%; Pred. No. 1.7e-17;
Matches 172; Conservative 103; Mismatches 302; Indels 255; Gaps 37;

QY 9 RINTAATVLAALSSVFAAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTKIPA 68
DB 2 RITTLASVTPCLGFS--ASSIAAEDVMIVSASGY-----EK----- 37
QY 69 SLREIPQSVSIITNQVKDRNVDTPDQARKTGLRLVLS--NDDGRSSVYARGY--EYSE 124
DB 38 KLTWAAASVSVISEEELQSSQYHDLAELARSVEGVDSGTGGLGISIRGNPASYTL 97
QY 125 YNIDGL-----PAQMOSIN-GTLPNLPADFVRVEMRGPGLFDSSGEMGGIVNLVR 174
DB 98 ILIDGVRQGGSSDVTPTNGFSAMNTGFMPPLAAIERIEVIRGPMSTLYGSDAMGGVNIIT 157
QY 175 KRPTKATQGHAAAG-----FGTHQYK-----AEDVSGSLNSDGSVRGR-----V 215
DB 158 RKNADKWLSSVYNAGLNQLQESNKNWNSQFNFWSGGLVDDSVSLQVRGSTQQRQGSVTS 217
QY 216 MAQTVGAS--PRPAEKNRHETFYAAADWDINPDTVLGAGLYQOORHLAPYNGLPADANNK 274
DB 218 LSDTAGTRIPYPTESQYVN--LGARLDWKASEQDVL----- 251
QY 275 LPSLPQHFVFGADWNKFKMNSHDVFADLKHVFGNGYGVKGMRYSDRSDADSNVAFAGSKL 334
DB 252 -----W--FDMDT-----TRQRYDNRDGO-----LGSALT 273
QY 335 GSKTPAGRPGCNTADDKACAVGLGTEIKQALAFDASYSPFRLGNTANEFVIGADYRFR 394
DB 274 G-----GYDRTLRYERNKISAGYDGTFTFTGWS-----YLNW 306
QY 395 RSTNEQGRITLYA-----RGGLALNEFRSIPQVDLIANA-RKGVGRYSHTVA----- 440
DB 307 NETENKRELVRSLKXDKWGLA-GQPRELKESSLNLSLLTLPGLSHLTVGGEFQSS 365
QY 441 -----TENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGESGKTLHKASK 487
DB 366 SMKDGVVLASTGETFRQKSWSPAEDEWHLTDALALTAGSRYEHHQFGGH----- 416
QY 488 TKFTGYAGAVYDLNDNNSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQ 531
```



```

Db      417  --FSPRAVLVWDADADMTLKGVTGKARWQQLHKGISVSGQGKTNLGNPDLKPEE 474
Qy      532  GNQFEVGYKGSYMD--RLNARVSYFMKDKNAAAPLNPNKKTTRYAALGKRVMEGYETE 589
Db      475  SVSEYAGV---YVDNAGLANANTGFWTDSNKIVSISINDNTNNSYNSGKARLHG--E 529
Qy      590  ISGAVTPKQO---IHAGSYLHSQIKTASNSRDDGIFLIMPHSANLMTTYQVTPLELT 645
Db      530  FAGTL-PLWSEVDYTLISINTYWTRESEQRDGNK--GAPLSTYPEHMYNAKLMQITEEVAS 586
Qy      646  GCGV-----NMSGITSSA---GMHAGYATFDMAAYRFTPKLQINAD 688
Db      587  WLGAARYGKTPRTQNYSSLSAVQKYYDEKGYLKAWTVVDXGLSMKMTDALTLNAAYN 646
Qy      689  NIFNRHY-----YARVGANT-FNIPGSEPTWTANLARYSF 722
Db      647  NLNKKYSDVSLXSAGKSTLYAGDYFQTSSTTGVIYF--ERYMWSLNTQF 696

```

RESULT 30

```

AA72914
ID  AA72914 standard; Protein; 725 AA.
XX

```

```

AC  AA72914;
XX

```

```

DT  13-JUN-2001 (first entry)
XX

```

```

DE  E. coli ironNc extracytoplasmic protein fragment.
XX

```

```

KW  ironNc; extracytoplasmic protein; immunogen; vaccine; E1; UTI;
KW  immunotherapy; extraintestinal infection; urinary tract infection;
KW  meningitis; pneumonia; intra-abdominal infection; antibiotic.
XX

```

```

OS  Escherichia coli.
XX

```

```

FH  Key Location/Qualifiers
FT  Peptide 1..24 /label= Signal_peptide
FT  Protein 25..725 /label= Mature_E. coli_ironNc_extracytoplasmic_protein

```

```

FT  Misc-difference 120..121 /note= "Encoded by GTG CGT TAT"
FT  FT /note= "134..136 /note= "Encoded by AAC TGG"

```

```

PN  WO200121636-A1.
XX

```

```

PD  29-MAR-2001.
XX

```

```

PF  22-SEP-2000; 2000WO-US26117.
XX

```

```

PR  22-SEP-1999; 99US-0155621.
XX

```

```

PA  (UTM) UNIV NEW YORK STATE RES FOUND.
XX

```

```

PI  Russo T, Carlino U;
XX

```

```

DR  MPI: 2001-244936/25.
XX

```

```

DR  N-PSDB; AAD03054.
XX

```

```

PT  Novel isolated ironNc polynucleotide from extraintestinal isolate of
PT  Escherichia coli useful as vaccine for treating or preventing
PT  extraintestinal infections caused by extraintestinal pathogenic
PT  Escherichia coli -
XX

```

```

PS  Claim 9; Page 39-41; 4dp; English.
XX

```

```

CC  The invention relates to the identification of ironNc gene,
CC  from an extraintestinal isolate of Escherichia coli. This
CC  gene is expressed in increased amounts in human urine and is
CC  identified by transposon (trnpho) mutagenesis. ironNc gene
CC  encodes an extracytoplasmic protein. This gene can be used as

```

```

CC  an immunogen in vaccine formulations. The recombinant vector
CC  comprising nucleotide sequence encoding one or more antigenic
CC  epitope of ironNc is useful for diagnostic and immunotherapeutic
CC  purposes. The ironNc antigenic peptide is useful for treating or
CC  preventing extraintestinal infections (EIs) caused by extra-
CC  intestinal pathogenic E. coli (ExPEC). The EIs include urinary
CC  tract infection (UTI), meningitis, intra-abdominal infection
CC  and pneumonia.
CC  The present sequence is E. coli ironNc extracytoplasmic protein.
SQ

```

```

Sequence 725 AA.

```

```

Query Match 7.2%; Score 273.5; DB 22; Length 725;

```

```

Best Local Similarity 19.8%; Pred. No. 1.5e-15;
Matches 167; Conservative 127; Mismatches 304; Indels 247; Gaps 40;

```

```

Qy      5  MSYFRINMATAATVLAASSVPAQCPADLETHIKQGRSYNAIVTEKNDYSFAVTGT 64
Db      1  MRINKLMSLTVLVLGINSQVSAKYS-----DDNDETLVVA 39

```

```

Qy      65  KIPASLREIPOSVSIITNQOVKDR-NVDTFDLARKTPGLRVLSN-----DDGRSSVY 116
Db      40  TAEQVLAKQRP-GVSVITSEDIKTPPVNDLSDIIRMPGVNLTGNSASGTRGNRQIDIR 98

```

```

Qy      117  ARGYEYSEYNDIGLPAQMS-----INGTLPLNFAFDVVEVMRQPSGLPDS 162
Db      99  GMGPENTLILIDGVPVTSRNSVYSWRGERDTRGDTRWVP--EQVERIEVINGPAARYG 157

```

```

Qy      163  SGEMGIVNLVLRKPTKAAQGHAAAGFTHKQYKAA-----DVSGSLNSDG---SV 211
Db      158  SGAAAGVAVNIITKRPPTND--HGSLSLYTNQPESSSEGAIRAFSLSPGLADALTTL 215

```

```

Qy      212  RGRV-----MAQTVASPPAPAEKNNRHETFYAAADMDINPDVLG--AGLYQQRH 260
Db      216  YGNLAKTIDADSDINDINSPVGTAKAAGHEGVRNMDINQVSMKLNPOQILPEVGYSRQ-- 273

```

```

Qy      261  LAFYNGLPADANNKLPSPLOHVV--GADMNKFKMNSHDVFDLKYFPNG--GYGRV 314
Db      274  ---NIVADDTQNSSSSAVTESLAKSGKETNRLYRQNYGI---TH--NGIWMQGSRF 322

```

```

Qy      315  GMRYSRDLADSNVAFGSLGKMTKTPAGRGCTADDKACAVALGIGTIRKQALAFDASYR 374
Db      323  GYYEYKTN-----NTMNBGLSGGSGRILAGK-----FTNRLS 358

```

```

Qy      375  PPLIGNTAN-----EFVIGADYNRFRSTNEQRTTLVARGGLALNFPSPQVDLI 425
Db      359  SWRTSGELNIPLVNMTDQTLTVGAEMNRDKLDPSST--LTVND-RDI----- 404

```

```

Qy      426  ANARKGVGYSHVATENLDEF-GIVGKSTFPADGLSLIGGR----- 468
Db      405  ---SGISGSADRSKNSQSIALYIEDNIEVPETNIIIGLRFDYLDSDGSGFSPSLN 460

```

```

Qy      469  ---LGHY-KISSGSGKT-----LHKASKTKFTGVA-----GAYYDLNDNSLYLSLQ 512
Db      461  LSGELDYFKYKAVAGARTKAPNLQSSSE---GYLYSKGNGCPDINSGCCYLLG--- 513

```

```

Qy      513  LYTPTNLADGKLLKPRQGNQFEVGYKGSYMDRLNARVSYFMKDKNAAAPLNPNKK 572
Db      514  ---NKDLN-----PEISVNKEIIGLETFWED--YHSAVTFIRNDYQNKIYAGDVICQ 560

```

```

Qy      573  TRYAL-----GKRVMEGVETEISGAVTPK---WQHAGSYLHSQIKTASNSRDDGI 622
Db      561  TSGAVILKMQNGKALVDIGIEASMSFLVYKRLMNTNATM-----MISSEQDTGN 613

```

```

Qy      623  FL-LMPKHSAN-----LWTTY-----QVTPLELTGGGVNANSGITTS 657
Db      614  PLSVIPKYTIINSLMTITQAFSASFMTLYGRQKPRTHAETRESEPT--GGL----- 663

```

```

Qy      658  SAGMHAGGATDAMAAYRFTPKLKQIINADNIPNRYAVARQANTFNIIPGSEPTWTAN 717
Db      664  SKGEIAYSLVGTNPNYINKNRLNVGVSINLKKQIFRSSEGANVTNPPG--RAYYAG 720

```

```

Qy      718  LRYSF 722

```


CC pneumoniae, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acid can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 813 AA;

Query Match 6.3%; Score 237.5; DB 22; Length 813;
 Best Local Similarity 20.6%; Pred. No. 3.5e-12;
 Matches 171; Conservative 107; Mismatches 301; Indels 253; Gaps 38;

QY 19 AALSSVFAAQTADLETTHIKGGRSYNAIVTEKNQDYSSPAVVGKIP-ASLRRIPO-S 76
 DB 107 AAVPATOARSEPLDHEQMETTSRTSSDVS-----ATROSTVIEHQLLEBRQGS 157
 QY 77 VSIITNQYKDNVDTFQDLARKTPGLRVLSNDGRSSVYARGEYSE-----YNID 128
 DB 158 DSLAT-----VLAKVPGM-----SDSSRTIT-----ELGQTLRGSMVWMD 195
 QY 129 GLPAQM-OSINGTLPNL--PADRVFVMKPGSLPDSSEMGIVNLVKRPKAFQGH 185
 DB 196 GYPLMTNRSSNNLANIDALIERIEIVIRGSSAIY-GSGATGIIIS-ITTRP----- 245
 QY 196 AAGFGTHKQYKAADVSGS-----LNSDGSVVG--RWAQVYGAAPRAEKNRHETVY 237
 DB 246 AGG-----ENRAETRLSATSPYLRIGSDG-LGQFOOYFAGSGALDYSPDFGTRH----- 295
 QY 238 AAADWDINPDYVL--GAGYLYQORHLAPYNN-----GLPADANN----- 273
 DB 296 VGASVDHAGDRAPERSQDGLPDSN--VYNIGCKIGLAIIDENQROVALSHYDARQDPT 352
 QY 274 -----KLIP--SLFOHVFVGADNMKFKMNSHDFADLKHYFFGNGGKVGKGRYSDRDA 323
 DB 353 YATDPRVALPGSVYANAIKGHELDE-QNRIRNTLANLEY----- 392
 QY 324 DSNVAFAGSKLGMK-----TPAGRPGCTADADKACAVGLGEITKOKALAPDASY 373
 DB 393 -ENLDLIGSRISAQLYYRDYFTRFTFPDARAVSTRG-----GNVDQIMQNSVFGSRLT 445
 QY 374 RPPRLGNTAN-EFVIGADYNRFRS--TNEQGRITLYARGGLALNFRSIPVDLITANAR 429
 DB 446 LATPLGESGNTLVWGGDINQGRSDMPDLVFDPAADAGGLVFDKIGLITTPPLRTS 505
 QY 430 KCV-----RGYSHVTATENLDEFGIYKSKTFHPADGLSLIGGRLGHYKI 474
 DB 506 AGAFPAQLQHRFDEHWSIDGGLRYEYSTAEFDPIPLSES--KAASPVYKGGDLIDYDAN 562
 QY 475 EGGEGKTLHAKSKTKRTGYAGAVDYLDNNNSLYLSL-----QYTPQTNIDAG 524
 DB 563 LSN-----LGIYVSPVACQEIYASFSQGFOLDVGIQLRNARGFDIS 606
 QY 525 KLLKPRQGNQFVGYKGYMDRLNARVSFYRMKDKNAAPLPPNNKRTRYALGKRV 584
 DB 607 SNIPEPKTNNYELGMRGALCGNTLSGLAFYTTSSKLGDVOSFNGILIR-----TKERY 662
 QY 585 GVE-----TEISGAVTP--KWOIHAGY-----SYLHSQIKTAS 615
 DB 663 GVEASADWLSDDEVMGAGSATMWRGRERKPDGKQDMTGVRPPLKLAVALYQKRDADW 722
 QY 616 NSRDSOIFILMPKHSANMTTYQVTEPLTIGGIVNMSGITSSAGNHAGYATFDMAAY 675
 DB 723 NNRLQATFPDSKDY-----RLDGVESFGRQVSTYTTTVDLVSOY 761

QY 676 RFTPKIKIQINADNIFNRHYARV-----GGANTFNIPGSEPTWTANLYSF 722
 DB 762 RTTPDQSLIGIQLNFNRDYPLYSQLLNNNNNTSHLPAPGTVLASYYTHW 813

RESULT 33

ID AAY38830 standard; Protein; 211 AA.

XX AAY38830;

XX 08-OCT-1999 (first entry)

XX Neisseria meningitidis antigen encoded by a partial ORF23.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 XX treatment; Neisseria infection; meningitis; septicemia; gonorrhea.

OS Neisseria meningitidis.

PN W0924578-A2.

PD 20-MAY-1999.

PF 09-OCT-1998; 98WO-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

PA (CHIR-) CHIRON SPA.

PI Grandi G, Maignani V, Pizza M, Rapuoli R, Scarlato V;

DR WPI: 1999-327407/27.

DR N-PDB: AA212258.

XX Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for
 XX diagnosis, treatment and prevention of infection

PS Claim 4; Page 379; 524pp; English.

XX Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*
 XX and *N. gonorrhoeae* antigenic proteins. They are encoded by open
 XX reading frames (ORFs) AA211972-212358. The antigenic proteins,
 XX their fragments, their nucleic acids and antibodies are used for
 XX diagnosis, prevention (as vaccines) or treatment of *Neisseria*
 XX infections, such as meningitis, septicemia and gonorrhea. Both
 XX organisms are closely related. Fragments of the nucleic acids
 XX are useful as hybridisation probes and antisense reagents.

XX Sequence 211 AA;

Query Match 6.2%; Score 235.5; DB 20; Length 211;
 Best Local Similarity 31.5%; Pred. No. 6.2e-13;
 Matches 64; Conservative 34; Mismatches 94; Indels 11; Gaps 3;

QY 111 GRSSVYARGYSEYNIDGLPAMQOSINGTLPLNLFAPDRVEYVMRGPGLPDSGEMGIV 170

DB 1 GYNYLFAKRSRIANQIINIPVADALADIGNANTAYEVEYVRYGVAAGLDGTGSESAIV 60

QY 171 NLVRKRPYKAFQGHAAAGFTKQYKAADVSGSINSDGVSRYVAAQTGASPREAKN 230

DB 61 NLVRKRLTRKPLFEVRAEAGNKRHFGLDADVSGSLNTEKXLLAGRLVSTGRGDSWRERR 120

QY 231 NNEHETFYAADDINDPDYTLGAGYLYQGRHL--APYNGLPADANKPLSLQHFVVGAD 287

DB 121 SRXAEIYGLIEVDIAQFVRVHAXMDYQAQKERTADAPLSYAVVDSOGYATAFGPKDNPATN 180

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
 XX WPI, 2001-550253/61.
 XX
 XX A library of DNA fragments of Escherichia coli strains for the
 PT phylogenic determination of a given strain comprises polynucleotides of
 PT nature B2/D+ A- -
 XX
 XX Example 6; Fig 6; 646pp; English.
 XX
 CC The invention relates to a library of DNA fragments of Escherichia coli
 CC strains comprising polynucleotides (ABA8577-ABA8729 and ABA89533)
 CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
 CC antibacterial and immunosuppressive activity as part of pharmaceutical
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli
 CC infections. The polypeptides are useful for determining the phylogenic
 CC group of a given E. coli strain. These polypeptides can detect and treat
 CC an undesired development of E. coli, particularly an extra-intestinal
 CC infection that include systemic and non-diarrhoeal infections such as
 CC septicemia, pyelonephritis and meningitis this is particularly
 CC advantageous as bacterial resistance is increasing with the more
 CC frequent use of broad spectrum antibiotics.
 CC
 CC Sequence 682 AA:
 SQ
 Query Match 5 9%; Score 222.5; DB 22; Length 682;
 Best Local Similarity 20.1%; Pred. No. 6, 2e-11;
 Matches 159; Conservative 116; Mismatches 313; Indels 203; Gaps 33;
 QY 15 ATVLAALSSVFAAQTADLETYHKGSRYSNAIVTEKNDYSSFAVTGKLPASIREIP 74
 DB 12 AVMGSAVISSGYASSPKKEDTL-----VVTASGFTQOIRNAP 48
 QY 75 OSVSIITNOQVKRNVYDFDQARKTPGLRVLSNDGRSSVYARGY- EYSEYNIDGLPA 132
 DB 49 ASVSVTSEQLKKPVSVDLDAVKDVEGISITGNE-KPDISIRGLSGDYTLTVDG--- 104
 QY 133 OMOSIN-----GTLPLPAFDRVEVWRGSPGLPDSGSENGIYNLVKRPPTA 180
 DB 105 RROSGEBSRPNGSGGEAFPIPPVEAIEIEVIRGPMSSLYGSDAIGVYINITTK-PVNN 163
 QY 181 FQGHAAAGFG---THKOYKAADVSGSLNSDGSVGRVMAQGVASPPAEKNNHETP 236
 DB 164 QTWGVTLGIGGIIQEHGK-----GNSTTNDPYLSGPIKDKLGIQ----- 204
 QY 237 YAAADWDINDPVTLYGAGLYYQQRHLAPYNGLPADANNKLPISLPQHVFGADWNKFRPNSH 296
 DB 205 -----LYGGMNYRKED-SISQGTAPKDKNITATIQ--FTPTESQK----- 243
 QY 297 DVFAADIKHIFGNGYGVKGMRYSDRADNSVYAFASGKLGKTP-AGRPQCNATADDAACAV 355
 DB 244 -FFE-----YKKNQVHTLTPGESLIDMT-MRGVNLKQPNKRETHNSRSHVAAW 291
 QY 356 GLGTEI-----KOKALAFDASY-SRPFRLGNTANFVIGADYNNF 394
 DB 297 NAGGELHBEIAYQEKVIREVSKGKKDKYKMMDLNYESKPEITNT---IIDAVYTF 347
 QY 395 RSTNEQGRITLYARGGLANFRSIPQVDLIANARKGVGYSHTVATE--NLDEFGIYK 452
 DB 348 LPEN-----VLTIGG---QFOHAEIRDSATGK-----TBTQSVSIRKQAVPIE 390
 QY 453 STFHADGSLIGGRLGHYKTESGEGKTLHRAKSKTFPGYAG---AYDLNDNNSLYL 508
 DB 391 NEYAATDSLATLGGRLDNHEI-----YSSYNNPRLYAYYNLDNLTLLKG 435
 QY 509 STSOLY-TPQTNLDADG-----KLTKPQGNQPEVGYKSYMDRLNARV 552
 DB 436 GIAKAFRASTIEVSPFGFTLTOGGASIMYGNRDLKPEISVTEIIGIYS-NDSGFSASA 494
 QY 553 SFYRMKDKNAAPLNNKKT-----RYAALGKRMEGVETEISGAVTPKQJHAGYS 605

DB 495 TLENTDFKNKLTSTYDGTADPYTGLNTFIYDVNGEANINGVELAQIPIYDKMHSANYT 554
 QY 606 YLHSLKTAQNSNRDDGIFP-----LMPKHSANLMTTYGVTELLTGGVNMMSGITSSA 659
 DB 555 FTDSRRKSDDESL-NKSKLSKGEPLERTPRHANAKLEWYTDITPSSLN-YTGQIWA 612
 QY 660 GHHAG-----GYATPDMAAAYRFTPKLQINADNIFNRHYARVVGANTFNIPGSE 711
 DB 613 AQRNQAKVPRVANGFTSMIDIGLNTYQILPDTLLNFAVLANTDKRSEDIDTIDGMWQDEGR 672
 QY 712 RTWTANLRYSF 722
 DB 673 RYW-ANVRVSF 682
 RESULT 36
 AA80096
 ID AA80096 standard; Protein; 785 AA.
 XX
 AC AA80096;
 XX
 DT 16-MAY-2000 (first entry)
 XX
 DE TonB dependent receptor fccL amino acid sequence SEQ ID NO:25.
 XX
 KM Bacterium 2412.1; maize; detoxification; degradation; carcinogen;
 KM funonisin catabolic gene cluster; antifungal; mycotoxin.
 OS Bacterium 2412.1.
 PN WO200004158-A2.
 PD 27-JAN-2000.
 XX
 PF 14-JUL-1999; 99WO-US15837.
 XX
 PR 15-JUL-1998; 98US-0092953.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 PI Duvick JP, Maddox J, Gilliam J;
 DR WPI: 2000-171270/15.
 DR N-PsDB; AA291265.
 PT New isolated nucleic acid molecules which encode fumonisin degradative
 PT enzymes and transporters, used for detoxifying fumonisin or related
 PT mycotoxin or reducing the pathogenicity of fungi producing fumonisin -
 PS Claim 1, Page 148-151; 181pp; English.
 XX
 CC The present invention describes isolated nucleic acid molecules which
 CC encode fumonisin degradative enzymes and transporters from a
 CC Bacterium 2412.1 funonisin catabolic gene cluster. The Bacterium 2412.1
 CC was isolated from maize. The nucleic acid molecules can be used for
 CC producing an enzyme for detoxifying fumonisin or a structurally related
 CC mycotoxin, e.g. in harvested grain, or processed grain which is to be
 CC used as animal feed or silage. Ruminal microorganisms transformed with
 CC the nucleic acid molecules can be used as probiotic compositions or as
 CC feed inoculant compositions. Plants transformed with the nucleic acid
 CC molecules can degrade or transport fumonisin and can reduce the
 CC pathogenicity of a fungus producing fumonisin. The plants may be e.g.
 CC maize, sorghum, wheat, tomato, soybean, alfalfa, sunflower, Brassica,
 CC cotton or rice. The products can also be used as detection reagents for
 CC fumonisin and related compounds. AA291253 represents the Bacterium
 CC 2412.1 funonisin catabolic gene cluster, and AA291254 to AA291273
 CC represent isolated nucleic acids from Bacterium 2412.1 which encode
 CC AA80085 to AA80104, AA291286 to AA291291 represent sequence used in the
 CC exemplification of the present invention. AA291274 to AA291285 are given
 CC in the sequence listing from the present specification but are not
 XX specifically mentioned further.


```

Query March 5.4%; Score 205; DB 21; Length 187;
Best Local Similarity 27.6%; Pred.No.3,1e-10;
Matches 51; Conservative 43; Mismatches 71; Indels 20; Gaps 4

OY 8 PRINNTAATVLAALSS-----VPAAGTADLETHIKGQRSYNAIV--TEKNGDYSSF 58
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 6 FSLKILITWMLSLAYGGSFADGVVPVSDGNTSLDITVNVARS--HALLGKTEKTRSTYID 62
   | : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

OY 59 AATVGTPLPASIREIPQSVSIITNQGVKRNVDTPQLARKTEPLGRVLSLNDSDRSSVYAR 118
   :: : : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 63 RMTSTATGRIAGKDPQSVSVITRSLRDDKAHTTEEMAKNTTGGVNVVRDSSGQTRPLSR 122

```

```

79 IITNOQVXDRNDVTFPDQARKTPGLRLVLSNDDGRSSVYARGYESEYN-----ID 128
::: ::::: :::::

```


Db 139 VIRREDAKTATMREVLNRIPIGVSAENNGTSGHDLAMNFGIRGLNPLRLASRSTVLMD 198
Qy 129 GLPA-----QMOSINGTLNLPFAFDRVEVMRPGSLFDSSGEMGIVNLVRKRPTKA 180
Db 199 GIPVPFAPYQPOLSLAPVSLGNM---DAIDVVRGGAVRYGQPSGVGVNFV-----TRA 251
Qy 181 FQGHAAAGFGTHQYKAAADYSGSLNSDGSVRGRVMAQTVGNSPPAEKNRHEH----- 235
Db 252 IP-----QDFGIEAGVEGQLSPITSS-----ONNPKETHNLMLV 283
Qy 236-----FYA---AADWDIN-----PDTVLGAGVLYQORH-----LAPYNG---L 267
Db 284 GGTADNGFGTALLYSGRGSDWREHSATRIDDLMLKSKYAPDEVHTFNSLLQYIDGADM 343
Qy 268 PA-----DANNKLPSLPQHFVVGADNKNFKQNS--HDVFADLKHFGNGGYKGVKGRY 318
Db 344 FGLSLRADYDADRQWQSTRPYDRF---NGRRKLASLGIFQFOPDSQHKFNIQGFYTTTLR- 398
Qy 319 SDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPERL 378
Db 399-----SGLEQGRKITL-----SPRNYWVRGIEPRYSQIFMI 430
Qy 379 GNTANFVIGADYNFRSTNEQRTTLVARGGLALNEFRSIPQVDLIANARKGVRGYSH 438
Db 431 GPSAHEVGVRYLN-ESTHEMYRYTATSSGQLPSG---SSP---YDRDTRSGTEAHAW- 482
Qy 439 VATENLDFYIGKSTFHPADGLSLIGGRGLGHYKIESGEGK---TLHKASKTKFTGYA 494
Db 483 ---YLDKIDIGNWTIP-----GMRPEH---IESVQNNAITGTHEEVSYNAPLAL 528
Qy 495 GAVYDLNNSLYL-----SLSOLYTPQTNLDADGKLLKPRQNGQEVGKSGSYMDRLNA 550
Db 529 NVLYHLTDSWNLNANTEGSGFTVQVSOIGKAVQSGNVPEPEKARTWELGTR--YDDGALTA 586
Qy 551 RVSEYVRMKDKNAAAPLNPNKKTRYAALGKRVMEGVETEIS---GAVTP---KWOIHAGY 604
Db 587 ENGLFLIFNN---QYDSNQNTDVTARGKTRHTGLETQARYDLTGLTFLTLNVSIIYASY 643
Qy 605 SYLHSGIKTASNSRDDGIFLLPKHSANL-----WTTYQVTPPELTIGGGVNMAMSGIT 656
Db 644 AYVNAIREKGTGYN-LVPFSPKHKGLGVDPKGNW-TFNLSNDFOSSQPADNANTVK 701
Qy 657 SSAGHAG---GYATFDMAAARFTPK---LKLQINADNIFNRHYVARVGGANTFNI-PG 709
Db 702 ESADGSTGRIPGFMWGARVAYDFGQWADNLAFGVKNIFPDQYFIRSYDDNNKGIYAG 761
Qy 710 SERT 713
Db 762 QPRT 765
RESULT 40
ABG17860
ID ABG17860 standard; Protein; 454 AA.
XX
AC ABG17860;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17851.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS82047.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 48219; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 454 AA;
Query Match 5.2%; Score 196; DB 22; Length 454;
Best Local Similarity 24.7%; Pred. No. 8.4e-09;
Matches 74; Conservative 52; Mismatches 127; Indels 46; Gaps 11;
Qy 10 INMTAATVLAALSSVFAAQADLETVHIKQORSYNAIVTEKNGDYSSFAVTGPKIPAS 69
Db 55 VAVVAVATAVSGMSVYAQAAPVEPKEDTITVTAAPAPQESAMGPAATIAARQSATGKTDT 114
Qy 70 LREIPQSVSIITNOQVDRNVDTFDQLARKTPGLRVLSNDDGRSSVY-----ARGY 120
Db 115 IOKVPQSVISVVTABEMALHQPKSVEALSYPGVSV--GTRGASNTYDHLIRGPAASGQ 172
Qy 121 EYSEYNIDGLPAQMOSINGTLNLPFAFDRVEVMRPGSLFDSSGEMGIVNLVRKRPT-- 178
Db 173 SONNY-LNGLKLGQNFYNDVIDPYMLERAEMRGPVSVLYGKSNPGGLLNWVSKRPITE 231
Qy 179 --KAFQGHAAAGFGTHQYKAAEADVSGSLNSDG--SVGRVMAQTVGASP----- 224
Db 232 PLKEVQPKA----GTDSLFTQGFDFSDSLDDDGVSYSYRLTGLARSANAQKQSGSEQRV 287
Qy 225 -----RPAEKNRHEHETFYAAADWDINDPDTVLGAGVLYQORHLPY-NG--L 274
Db 288 APAFTWRPDDKTN-----PTFLUSYFQNEPDTGY-YGWLPEKGTVEPLPKGLPTDFN 341
RESULT 41
AAB40909
ID AAB40909 standard; Protein; 134 AA.
XX
AC AAB40909;
XX
DT 08-FEB-2001 (first entry)
XX

CC specification, but was obtained in electronic format directly from WIPO.
CC at ftp.wipo.int/pub/published/pct/sequences.

```

SQ Sequence 134 AA;
Query Match          5.2%; Score 195; DB 23; Length 134;
Best Local Similarity 39.0%; Pred. No. 1.5e-09;
Matches 46; Conservative 20; Mismatches 38; Indels 14; Gaps 3;

QY 457 PADGLSLIGGRGLGHYKTESGKTT-LHKASKTK-----FTGVAGAVYDLNDNNS 505
      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 5 PTDDLVLIGTRVSNF---SGTDNTDFYDPKADNRLTYRQTGVVTPYAGIVYDLNDIWS 61

QY 506 LYLSELQYTPQTNLDADGKLLKPRQGNQFVGVYKGSYMDRLNARVSFYRMKDKNAA 563
      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :
Db 62 VYTSVTKYKSPNSKADPKLLDIEGTYTBAGLKAAFFDGLRLNAPSPAPRIEODNVA 119
      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : :

```

RESULT 43	
ABG29752	
ID	ABG29752 standard; Protein; 340 AA.
XX	
AC	ABG29752;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #29743.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
RT	Drmanac RT, Liu C, Tang YT;
PI	
XX	
DR	WPI; 2001-639362/73.
DR	N-PSDB; AAS93939.
XX	
PT	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
XX	Claim 20; SEQ ID No 60111; 103bp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques for restoring normal activity of (II) or to treat disease states involving (III). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (I) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ARG000010-ARG30377 represent novel human

CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published pct sequences](http://pub.int/pub/published/pct/sequences).

```

SQ Sequence 340 AA;

Query Match 5.2%; Score 195; DB 22; Length 340;
Best Local Similarity 27.1%; Pred. No. 6.6e-09;
Matches 48; Conservative 26; Mismatches 69; Indels 34; Gaps 4;

Qy 528 KPROGNQFEVGYKGSYVDDRLNARVSFYRMKKKNAAPLNPNNKKTRYAALGKKRYMEGVE 587
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 56 KPQKANTSEIGTKWQVLDRLLTAAALFRTDIENEV----EQNDGDGYSYQGKKRVEGYE 111

Qy 588 TETSGAVTPKQWQHAGYSYHLHSQIKTASNRDDGIFLL--MPKHSANLWTTYQVTPELTI 645
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 112 ISVAGNITPAWQVIGGYTQQAQATIKNGKDVAQDGSSELPTYPEHAFTLWSQYQATDDISV 171

Qy 646 GGGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQINADNIFNRHYARVYGGG 702
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 172 GAGARYI-----GSMHKG-----SDCANNETYGFHVDGA 200

```

RESULT 44
AAY34494
ID AAY34494 standard; Protein: 833 AA.

AC	AA34494;
XX	
XX	25-AUG-1999 (first entry)
DT	
XX	
XX	Porphorymonas gingivalis protein PG35.
XX	
XX	Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW	vaccine; antigenic.
KW	
XX	
OS	Porphorymonas gingivalis.
XX	
XX	WQ929870-A1.
PN	
XX	
XX	17-JUN-1999.
PD	

XX	10-DEC-1998;	98WO-AU01023.
XX		
XX	04-AUG-1998;	98AU-0005028.
PR	10-DEC-1997;	97AU-0000839.
PR	31-DEC-1997;	97AU-0001182.
PR	10-JAN-1998;	98AU-0001546.
PR	10-MAR-1998;	98AU-0002264.
PR	09-APR-1998;	98AU-0002911.
PR	23-APR-1998;	98AU-0003328.
PR	05-MAY-1998;	98AU-0003338.
PR	22-MAY-1998;	98AU-0003654.
PR	29-JUL-1998;	98AU-0004947.

PA (CSLC-) CSL LTD.

xx Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
 PI Ross BC, Rothel LJ, Webb EA;
 xx WPI: 1999-385613/32.
 DR N-PSDB: AAX91712.
 DR

Antigenic Porphyromonas gingivalis peptides for preventing gingivitis

PS Claim 1: Page 474-475: 588pp: English:

AA AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
CC AAX34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC

CC activity with a vaccine mechanism of action. The PG polypeptides can be
 CC used as vaccines especially against Porphyromonas gingivalis. Probes can
 CC be used to detect Porphyromonas gingivalis in standard hybridisation
 CC assays. Porphyromonas gingivalis is involved in periodontal disease
 CC especially gingivitis.

XX Sequence 833 AA;

Query Match 5.1%; Score 193.5; DB 20; Length 833;
 Best Local Similarity 20.4%; Pred. No. 3.7e-08;
 Matches 142; Conservative 82; Mismatches 266; Indels 207; Gaps 32;

QY 34 EYVHIK--GQSYNAIVTEKNGDYSFAVTGKIPASLREIPQSVSIITNOQVQRND 91
 DB 88 EKVIHIEKGSRRHVDLYLTELIELSLDGVVVS--ANRNETFRQAPSLVTVLSPELFLKTNST 146
 QY 92 TPDQLARKTPGLRVLSN--DDGRSSVYARGYE--YSE-----YNIDGLPA 132
 DB 147 NLSQGLKFPQGLRVEDNCCQGFNOVRINGLEGAYSQIILDSHPITFSSLAGVYGLQMPA 206
 QY 133 QMOSINGTLPNLFAPFRVEMVRGPGSLFDSSGEMGIVNLVRRKPTK--AFQGHAAAGFG 190
 DB 207 NM-----IERVEYIRGGGSGALFGSNVGVGINVITEPLRNSAEISHSTMTFD 254
 QY 191 THKOYAEADVSGSLNSDGSVRGRVMAQTVGASPRPAEKNNRHETTYAADMWDINDPTVL 250
 DB 255 HAKGW-----GSFQNTQOFNGSMLTE-----DRKA 279
 QY 251 GAGYLKQQRHLAPYNLPPADANNKLPSPHNVGVGADMKFKKNSHDVPADLKHFGNGG 310
 DB 280 GV-MVFGQHNHYRGGQDIDGNFTFELPNLRNRL-----GFRSYKKTGL 321
 QY 311 YGVKGRYSDRDRDADSNYAFAAGSKLGMKTPAGRPGCNTADKACAVGLTEIKQKALAFDA 370
 DB 322 YSKATLEIYHSM--QERYRGGRDL-----NPPFAQIAEYLQHYINGSGFRFDQ 368
 QY 371 SYSRP--FRLGNTANE-----FVIGADY--NFRSTNEQRTTL--YARGIALNEFRS 418
 DB 369 GSGGKGFPSLYASQDVGRSRYGGDYENLNGAVQSGSTESDEYNDAPFALTSYGT 428
 QY 419 IPQVDLIANARKGVRSHTVATENDEFG---IYK-----STPHPA 468
 DB 429 YKGFEDL-----QGGMYRHTFG--ENWDFTGLEYIGQLDRSGYRPSKIDONTSTFSQY 482
 QY 459 DGLSLIGGRGLGHYKIESGEGKTLHKASKTKFTGYAGAYVD--LNDNNSLYLSOLYT 515
 DB 483 DQI-----EYKTE-----KLSALIGARLDYLLANDGKRYIDPLPIFS 520
 QY 516 PQTNLDADGKLLKPRQGNQFVGYK-----GSYMDRLNARVSPYRMDKNNAAP---L 566
 DB 521 PRANV-----RYPNPNKRLSPRLSYSEGFRAQYFDEDLHVELA-----GGTPISRVL 567
 QY 567 NPNNKKTIRVAALCK-----RVN--EGVETEISCAVTPKQOIHAGYISYLSQIK 612
 DB 568 SPMLKEERSRSISASPDYHRADEWFNIMGAEFSTFISNQFKPSDKVETTSDEKWIIR 627
 QY 613 TASNRRDDGIFILMPKHSANL--WTTYQVTPPELLIG 647
 DB 628 TIYNDKDG-----VSKYGVNLEGRIRIYNKSFDLQGG 660

RESULT 45
 AAY34368
 ID AAY34368 standard; Protein; 876 AA.
 XX AAY34368;
 AC AAY34368;
 XX
 XX 25-AUG-1999 (first entry)
 DT
 XX Porphyromonas gingivalis protein PG35.
 DE Porphyromonas gingivalis; PG, periodontal disease; gingivitis;
 XX vaccine; antigenic.
 KW

XX Porphyromonas gingivalis.

OS WO9292670-A1.

XX 17-JUN-1999.

XX 10-DEC-1998; 98WO-AU01023.

XX 04-AUG-1998; 98AU-0005028.

XX 10-DEC-1997; 97AU-0000839.

XX 31-DEC-1997; 97AU-0001182.

XX 30-JAN-1998; 98AU-0001546.

XX 10-MAR-1998; 98AU-0002264.

XX 09-APR-1998; 98AU-0002911.

XX 23-APR-1998; 98AU-0003128.

XX 05-MAY-1998; 98AU-0003338.

XX 22-MAY-1998; 98AU-0003654.

XX 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;

XX Ross BC, Rotnel LJ, Webb EA;

XX WPI; 1999-385613/32.

XX N-PSDB; AAX91586.

XX Antigenic Porphyromonas gingivalis peptides for preventing

XX gingivitis

PS Claim 1; Page 330-331; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic

XX Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34338 to

XX AAY34383. AAX91802 to AAX91989 represent PCR primers used in the

XX isolation of the PG polypeptides. The PG polypeptides have antibacterial

XX activity with a vaccine mechanism of action. The PG polypeptides can be

XX used as vaccines especially against Porphyromonas gingivalis. Probes can

XX be used to detect Porphyromonas gingivalis in standard hybridisation

XX assays. Porphyromonas gingivalis is involved in periodontal disease

XX especially gingivitis.

CC Sequence 876 AA;

Query Match 5.1%; Score 193.5; DB 20; Length 876;
 Best Local Similarity 20.4%; Pred. No. 4e-08;
 Matches 142; Conservative 82; Mismatches 266; Indels 207; Gaps 32;

QY 34 EYVHIK--GQSYNAIVTEKNGDYSFAVTGKIPASLREIPQSVSIITNOQVQRND 91
 DB 131 EKVIHIEKGSRRHVDLYLTELIELSLDGVVVS--ANRNETFRQAPSLVTVLSPELFLKTNST 189
 QY 92 TPDQLARKTPGLRVLSN--DDGRSSVYARGYE--YSE-----YNIDGLPA 132
 DB 190 NLSQGLKFPQGLRVEDNCCQGFNOVRINGLEGAYSQIILDSHPITFSSLAGVYGLQMPA 249
 QY 133 QMOSINGTLPNLFAPFRVEMVRGPGSLFDSSGEMGIVNLVRRKPTK--AFQGHAAAGFG 190
 DB 250 NM-----IERVEYIRGGGSGALFGSNVGVGINVITEPLRNSAEISHSTMTFD 297
 QY 191 THKOYAEADVSGSLNSDGSVRGRVMAQTVGASPRPAEKNNRHETTYAADMWDINDPTVL 250
 DB 298 HAKGW-----GSFQNTQOFNGSMLTE-----DRKA 322
 QY 251 GAGYLKQQRHLAPYNLPPADANNKLPSPHNVGVGADMKFKKNSHDVPADLKHFGNGG 310
 DB 323 GV-MVFGQHNHYRGGQDIDGNFTFELPNLRNRL-----GFRSYKKTGL 364
 QY 311 YGVKGRYSDRDRDADSNYAFAAGSKLGMKTPAGRPGCNTADKACAVGLTEIKQKALAFDA 370
 DB 365 YSKATLEIYHSM--QERYRGGRDL-----NPPFAQIAEYLQHYINGSGFRFDQ 411

Search completed: December 18, 2002, 06:40:21
Job time : 53.9653 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:40:49 ; Search time 16.862 Seconds
(without alignments)
1259.836 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776
Sequence: 1 MGQMSVPRINMTAATVLA.....NTFNIPGSERTANIRYSF 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfillset1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	296.5	7.9	696	1	US-08-765-081-5 Sequence 5, Appl
2	296.5	7.9	696	3	US-09-098-082-5 Sequence 5, Appl
3	296.5	7.9	696	5	PCT-US95-06994-5 Sequence 5, Appl
4	286	7.6	703	5	PCT-US95-06994-8 Sequence 8, Appl
5	273.5	7.2	725	4	US-09-668-113A-8 Sequence 6, Appl
6	264.5	7.0	718	5	PCT-US95-06994-6 Sequence 6, Appl
7	260	6.9	663	1	US-08-765-081-7 Sequence 7, Appl
8	260	6.9	663	3	US-09-098-082-7 Sequence 7, Appl
9	232.5	6.2	703	5	PCT-US95-06994-9 Sequence 9, Appl
10	212	5.6	652	1	US-08-765-081-6 Sequence 6, Appl
11	212	5.6	652	3	US-09-098-082-6 Sequence 6, Appl
12	186.5	4.9	791	3	US-08-537-361B-4 Sequence 4, Appl
13	186.5	4.9	791	4	US-08-817-707-4 Sequence 4, Appl
14	185	4.9	792	1	US-08-326-670A-2 Sequence 2, Appl
15	183	4.8	792	3	US-08-920-470A-2 Sequence 2, Appl
16	183	4.8	792	4	US-08-990-470A-2 Sequence 2, Appl
17	181	4.8	713	4	US-08-628-434-4 Sequence 4, Appl
18	180.5	4.8	718	5	PCT-US95-06994-7 Sequence 7, Appl
19	180	4.8	792	3	US-08-537-361B-2 Sequence 2, Appl
20	179	4.7	792	3	US-08-537-361B-6 Sequence 6, Appl
21	173	4.6	713	4	US-08-928-434-2 Sequence 2, Appl
22	173	4.6	915	4	US-08-613-009A-18 Sequence 18, Appl
23	173	4.6	915	4	US-08-778-570B-24 Sequence 24, Appl
24	173	4.6	915	4	US-09-059-584-24 Sequence 24, Appl
25	172	4.6	915	1	US-08-487-890A-96 Sequence 96, Appl
26	172	4.6	915	1	US-08-363-124A-2 Sequence 2, Appl
27	172	4.6	915	2	US-08-478-435-96 Sequence 96, Appl

28	172	4.6	915	2	US-08-337-483-96 Sequence 96, Appl
29	172	4.6	915	2	US-08-478-373-96 Sequence 96, Appl
30	172	4.6	915	3	US-08-474-671-96 Sequence 96, Appl
31	172	4.6	915	3	US-08-483-577A-96 Sequence 96, Appl
32	172	4.6	915	4	US-08-897-438-96 Sequence 96, Appl
33	172	4.6	915	4	US-08-637-654-96 Sequence 96, Appl
34	172	4.6	915	4	US-08-649-518-96 Sequence 96, Appl
35	171	4.5	909	2	US-08-363-124A-4 Sequence 16, Appl
36	169.5	4.5	790	4	US-08-817-707-8 Sequence 8, Appl
37	169.5	4.5	908	3	US-08-613-009A-16 Sequence 16, Appl
38	169.5	4.5	908	4	US-08-578-570B-22 Sequence 22, Appl
39	169.5	4.5	908	4	US-09-059-584-22 Sequence 22, Appl
40	167.5	4.4	908	1	US-08-487-890A-94 Sequence 94, Appl
41	167.5	4.4	908	2	US-08-478-435-94 Sequence 94, Appl
42	167.5	4.4	908	2	US-08-337-483-94 Sequence 94, Appl
43	167.5	4.4	908	2	US-08-478-373-94 Sequence 94, Appl
44	167.5	4.4	908	3	US-08-474-671-94 Sequence 94, Appl
45	167.5	4.4	908	3	US-08-483-577A-94 Sequence 94, Appl

ALIGNMENTS

RESULT 1
US-08-765-081-5
; Sequence 5, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.R., Vary Jr., J.C.
; TITLE OR INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 696 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-765-081-5
Query Match 7.9%; Score 296.5; DB 1; Length 696;
Best Local Similarity 21.3%; Pred. No. 1.1e-18;
Matches 176; Conservative 103; Mismatches 304; Indels 243; Gaps 39;
QY 9 RIMTAATVLAALSSVFAQQTADLTETHIKGGRSNAIVTEKNGYSSRAIVGVKIRP 68
DB 2 RITTLASVPIPLGFGS--ASSIAADVMIIVASGY-----EK----- 37


```

QY 69 SUREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRVLS--NDDGRSSVYARG--EYSE 124
DB 38 KLTNAASVSVISQEBLOSQYHDLAEALRSVEGVDSGTGKTGLEISIRGMPASVTL 97
QY 125 YNIDGL-----PAQMOSIN-GTLPNLFADRVEMRGPSGLFDSSGEMGIVNLVR 174
DB 98 ILIDGVRQGGSDVTPNGFSAMNTGMPPLAAIERIEVIRGPMSTLYGSDAMGVVNIIT 157
QY 175 KRPTKAFQGHAAAG-----FGTHKQYK-----AEADVSGSLNSDGSVRGR-----V 215
DB 158 RKNADKWLSSVNAGLNQSNNKGNSSQFNWSSGPLYDSDSVLQVRGSTQQRQSSVTS 217
QY 216 MAQTVGAS--PRPAEKNNRHETFYAAADW-----DI-----NPDTVLGA---G 253
DB 218 LSDTAGTRIPYPTESQYNYN--LGAFLDKWAKSEQDVLWFDMDTTRQYDNRDQGLSLTGG 275
QY 254 YLYQORHLAPYNGLPADANNKLPSPQHVFGADWNKFKMNSHDVPADLKHVFGNGGYK 313
DB 276 YDRTURY-----ERNKISAGYDHTFTFTGWSY-LNNWET-----ENKREL 316
QY 314 VGMRYSDRDADSNYAFAGSKLGMKTPAGRPCNTADDKACAVGLGTEIKOKALAFDASYS 373
DB 317 V-----RSVLKRDKWGL---AGQP-----RELKESNLILNSLLL 347
QY 374 RPFRLGNTANEFVIGADYNRFRSTNEOGRITLYARGGLALNEFRSIPQVDLIANARKGVR 433
DB 348 TPL-----GESHLVTVGG-----EFQSSSMKDGIVLSTG-- 377
QY 434 GYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGY 493
DB 378 -----ETFRQKSWSVFAEDEWHLTDALALTAGSYEHHEQFGH-----FSPR 420
QY 494 AGAVYDLNDNNSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQGNQFEV 537
DB 421 AYLVMDVADAWTLKGGVTTGYKAPRMGQLHKGISGVSGCKTNLLGNPDLPKEESVYEA 480
QY 538 GYKGSYMD--RLNARVSFYRMKDNAAAPLNPNKKTRYAALGKRVMEGVETEISGAVT 595
DB 481 GV---YYDNPAGLNANTGFMDFSNKIVSYINDNTNSVYNSGKARLHGV--EPAGTL- 534
QY 596 PKWQ-----IHAGSYLHSGIKTASNRDGIFFLLPKHSANLWYTYQVTPELTIGGV-- 649
DB 535 PLWSEBVTLSLNTWTRSQRDGNK--GAPLSYTPHEHVMNAKLWQITEEVASWLGARY 592
QY 650 -----NAMSGITSSA-----GMHAGGYATFDMAAAYRFTPKLKLQINADNIFNRH 694
DB 593 RGTFRFTQNYSSLASAVQKVVYDEKGEYLKAWTVVDAGLSWKTDTALLTNAAVNLLNKD 652
QY 695 Y-----YARVGGANT-FNIPGSBRTWTANLRYSF 722
DB 653 YSDVSLYSAGKSTLYAGDYFQTSSTGYVIP--ERNYMWLSLNYQF 696

```

RESULT 2

US-09-098-082-5

Sequence 5, Application US/09098082

Patent No. 6040421

GENERAL INFORMATION:

APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.

TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen, O'Connor, Johnson and Kindness

STREET: 2800 Pacific First Center, 1420 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

COMPUTER: IBM PC compatible/Pentium II

OPERATING SYSTEM: MS-Windows 95

SOFTWARE: Word for Windows-6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/098,082

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/765,081

FILING DATE: March 26, 1997

APPLICATION NUMBER: PCT/US95/06994

FILING DATE: June 7, 1995

APPLICATION NUMBER: US 08/265,714

FILING DATE: June 24, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: CHOR-1-12402

TELECOMMUNICATION INFORMATION:

TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)

TELEFAX: 1-206-224-0779

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 696 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-098-082-5

Query Match

Best Local Similarity 21.3%; Pred. No. 1.1e-18;

Matches 176; Conservative 103; Mismatches 304; Indels 243; Gaps 39;

QY 9 RINTAATVLAALSSVFAAQTADLETVHKGQSYNAIVTEKNGDYSSPAVTVGTKIPA 68

DB 2 RITTLASVVIPLGLFS--ASSIAAAEDVMIVASGY-----EK----- 37

QY 69 SLREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRVLS--NDDGRSSVYARG--EYSE 124

DB 38 KLTNAASVSVISQEBLOSQYHDLAEALRSVEGVDSGTGKTGLEISIRGMPASVTL 97

QY 125 YNIDGL-----PAQMOSIN-GTLPNLFADRVEMRGPSGLFDSSGEMGIVNLVR 174

DB 98 ILIDGVRQGGSDVTPNGFSAMNTGMPPLAAIERIEVIRGPMSTLYGSDAMGVVNIIT 157

QY 175 KRPTKAFQGHAAAG-----FGTHKQYK-----AEADVSGSLNSDGSVRGR-----V 215

DB 158 RKNADKWLSSVNAGLNQSNNKGNSSQFNWSSGPLYDSDSVLQVRGSTQQRQSSVTS 217

QY 216 MAQTVGAS--PRPAEKNNRHETFYAAADW-----DI-----NPDTVLGA---G 253

DB 218 LSDTAGTRIPYPTESQYNYN--LGAFLDKWAKSEQDVLWFDMDTTRQYDNRDQGLSLTGG 275

QY 254 YLYQORHLAPYNGLPADANNKLPSPQHVFGADWNKFKMNSHDVPADLKHVFGNGGYK 313

DB 276 YDRTURY-----ERNKISAGYDHTFTFTGWSY-LNNWET-----ENKREL 316

QY 314 VGMRYSDRDADSNYAFAGSKLGMKTPAGRPCNTADDKACAVGLGTEIKOKALAFDASYS 373

DB 317 V-----RSVLKRDKWGL---AGQP-----RELKESNLILNSLLL 347

QY 374 RPFRLGNTANEFVIGADYNRFRSTNEOGRITLYARGGLALNEFRSIPQVDLIANARKGVR 433

DB 348 TPL-----GESHLVTVGG-----EFQSSSMKDGIVLSTG-- 377

QY 434 GYSHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKFTGY 493

DB 378 -----ETFRQKSWSVFAEDEWHLTDALALTAGSYEHHEQFGH-----FSPR 420

QY 494 AGAVYDLNDNNSL-----YLS--LSQLYTPQTNLDADGKL-----LKPRQGNQFEV 537

DB 421 AYLVMDVADAWTLKGGVTTGYKAPRMGQLHKGISGVSGCKTNLLGNPDLPKEESVYEA 480

QY 538 GYKGSYMD--RLNARVSFYRMKDNAAAPLNPNKKTRYAALGKRVMEGVETEISGAVT 595

DB 481 GV---YYDNPAGLNANTGFMDFSNKIVSYINDNTNSVYNSGKARLHGV--EPAGTL- 534


```

QY 596 PKWQ-----IHGASYLHQLTASNSRQDGFILMPKRSALMTWYOTVELTIGGV-- 649
Db 535 PLMSEDVTLISLNTYWTNRSEQDGDGNK--GAPLSYTPPEHMVAQKINWQITEEVSASGTARY 5922
QY 650 -----NMSGITSSA-----GMHAGYATFDMAAARFPFKLQINDNPIFNRH 6944
Db 593 RGTBPRFTQNTSSLSAVOKKYDEKGEELKAMTVVDGLSKRMKDALTLLNAAVNNLNKD 6522
QY 695 Y-----YAEVGANT--FNIPGSERTTANLARYSF 722
Db 653 YSDVSLVSAGKSTLYAGDYFDTGSGSTTGIVLP--ERNYMSLNTQF 696

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	7.9%;	Score 296.5;	DB 5;	Length 696;
Best Local Similarity	21.3%;	Pred. No. 1.1e-18;		
Matches 176;	Conservative 103;	Mismatches 304;	Indels 243;	Gaps 39;

QY RINTAAVTLAASSVPAATATLETHIKGQSNVAITEKKGVDSSPAVVTGKTIPA 66
9 RINTAAVTLAASSVPAATATLETHIKGQSNVAITEKKGVDSSPAVVTGKTIPA 66
Db RITTLASVVICLGES--ASSIAAEEDMIVASGY-----EK----- 37
2 RITTLASVVICLGES--ASSIAAEEDMIVASGY-----EK----- 37
QY SLRELPQSVSITITNOQVKNRNVDTFDQILAKTPELRVLS--NDDGRSSVYARGY--EYSE 124
63 SLRELPQSVSITITNOQVKNRNVDTFDQILAKTPELRVLS--NDDGRSSVYARGY--EYSE 124
Db KLTAASAVSVISQEEIASSQHYDLAEKLRVSEGVAVDSEGTGKIGLIEIRKIPASVYL 97
38 KLTAASAVSVISQEEIASSQHYDLAEKLRVSEGVAVDSEGTGKIGLIEIRKIPASVYL 97

QY	125	YNIIGL-----PAMOMISIN-GTLPLNFAFRVWMEPGSLFDSSGEMGIVLVLR	174
Db	98	ILIDGVRQGGSSDVTPEMGFSANMTGMPPLAALIERIVIGMPSTLYGSDAMGAVNIIT	157
QY	175	KRPFAEQGHAAAG-----PGTHKOYK-----AEADVSGSLNSDGSVGR-----V	215
Db	158	RKNADKMILSSVNNGLNINQESNMKNSSQFNFWSGCLVDDSVSLQVRGSRQHGQGGSVTS	21.7
QY	216	MAQTVGAS-PPPAEKNRHHETFYAAADW-----DI-----NPDVLGA--G	253
Db	218	LSDDAGTRIPPEPTEGQYN--LGRILDMKASBODVLFMDTTRORYDNDDGOLGSLTIG	275
QY	254	YLVOGRHAPYNGPLRANANKLPBPLPHVAVGADVMNKFKNNSHDVFPADLLGHYRGNGGYCK	313.3
Db	276	YDRTLRY-----ERKNISIGYDHTTTFGTWSY-LNMNET-----ENKREL	316
QY	314	VGMRYSDRDADSNVAFAPGSKLGMKTPRPGCWATDDKCAVAGIGTEIKOKALAFDASYS	373
Db	317	V-----BSVLKRDKML---AGP-----RELKSNILINSLL	347
QY	374	RPFELGNTANEFVIGADYNNFRSTNEGRTLLYARGLALNEFRSIPQVDLIANKGVR	433
Db	348	TPL-----GESHLVTVG--EFQSSMKDGVVLASTG--	377
QY	434	GYSHTVATENIDEGITYGKSTFHPADGLSLGGRCILGHYKISBEGCKTLHKASKTYFTGY	493
Db	378	-----ETFRQKSWSVFAEDBWHHTDALATAGSRHYEHHEFOGHH-----FSPR	420
QY	494	AGAVYDLDNNSL-----YLS--LSQLTPTQNLDAQKL-----LKPRQGNQFEV	537
Db	421	AYLWMDVADAWTLKGAVTTGFKAPRMQOLHKGISGSGQKTYLLGNPDILKPESSVYEA	480
QY	538	GKYSYVMD--RLNARVSPYRMKDKNAAALPLNNPKKTRFAALGKRWMEGVFEISGAVT	595
Db	481	GV-----YYDNPRAGLNANTGTMDEPSNKIYVSINDYNTSVNSGKARLHV--EFACTL-	534
QY	596	PKWO-----IHAGSYLHSQIKTASNSRBDGI FILMPKHSANLMTTYQVTPELTIGGV-	649
Db	535	PLMSEVDVTLISNTYWTBRSBQRDGNK--GAPLSYTPBHMVNAKLNMQITREVASWIGARY	592
QY	650	-----NMSGITSSA-----GHHAGGYATFPDAAAYRTTPKIKLOINADNIPNRH	654
Db	593	RGTPTREFYONVSSLSAQQKYVDEKGEYLLKAMTVVAVAGLSWKMTDALLTNAAVNNLANKD	652
QY	695	Y-----YARVGANT-FNIPGSEBRTWANTLYSIF	722
Db	653	YSDVSLYSAGKSTLYADHYQTOGSSSTGYIP--ERYWMSLWYQF	696

```

1      RESULT 4
2      PCT-US95-06994-8
3      Sequence 8, Application PC/TUS9506994
4      GENERAL INFORMATION:
5      APPLICANT: Children's Hospital & Medical Center
6      APPLICANT: University of Washington
7      APPLICANT: Washington State University Research Foundation
8      APPLICANT: TARR, PHILIP I
9      APPLICANT: BIGGE, SIMA S
10     APPLICANT: BESSER, THOMAS E
11     APPLICANT: VARY JR, JAMES C
12     TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
13     NUMBER OF SEQUENCES: 14
14     CORRESPONDENCE ADDRESS:
15     ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINNESS PLLC
16     STREET: SUITE 2800, 1420 FIFTH AVENUE
17     CITY: SEATTLE
18     STATE: WASHINGTON
19     COUNTRY: USA
20     ZIP: WA 98101
21     COMPUTER READABLE FORM:
22     MEDIUM TYPE: Floppy disk
23     COMPUTER: IBM PC compatible
24     OPERATING SYSTEM: PC-DOS/MS-DOS

```


SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 703 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
DESCRIPTION: wherein "Xaa" residues represent gaps
DESCRIPTION: introduced to facilitate best alignment with
DESCRIPTION: SEQ ID NO:9.
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli O157:H7
STRAIN: 86-24 NALR
PCT-US95-06994-8

Query Match 7.6%; Score 286; DB 5; Length 703;
Best Local Similarity 20.4%; Pred. No. 1.1e-17;
Matches 171; Conservative 104; Mismatches 304; Indels 258; Gaps 38;
QY 9 RNNMTAATVLAAL---SSSVFAAQTADLTVHLKQORSNAIVTEKNGDYSSFAVTVGPK 65
DB 2 RITTLASVVPICLGFSSASSIXXXXXAAEDVMIVSAGY-----EK----- 42
QY 66 IPASLRIPQSVSIITNQVKDRNVDTPDLQARKTPGLRVLS--NDDGRSSVYARGY--E 121
DB 43 ---KLTNAAASVVISQEBLSQYHDLAELARSVEGVDVESGTGKTGGLGISIRGMPAS 99
QY 122 YSEYNIDGL-----PAQWQIN-GTLPNLFAPDRVEMRGSGLPDSSGEMGGIVN 171
DB 100 YTLILIDVROGGSSDVTNPGFSAMTGFMPPLAAIERIEVIRGPMSTLYGSDAMGGVYN 159
QY 172 LVRKRPKAFQGHAAAG-----FGTHQYK-----AEADVSGSLNSDGSVGRG--- 214
DB 160 IITRKNAKWLSSVAGLNLQSNKMGNSQFNFWSSGGLVDDSVLQVRGSGTQORQSS 219
QY 215 --VMAQTVGAS--PRAEKNRHETFYAAADWDINPDTVLGAGLYQORHLAPYGLPADA 271
DB 220 VTSLSDTAGTRIPYPTESQYN--LGARLDKASEQDVL----- 256
QY 272 NNKLPSLPOHVFVGDWPKFNKNSHDVFADLKHFGNGYGVKWMYSDRSDADSNVAFAG 331
DB 257 -----W--FDMDT-----TRQYDNRDQG-----LG 275
QY 332 SKLGMKTPAGRPCNTADDKACAVGLGTETIKOKALAFDASYSRPRRLGNTANEFVIGADY 391
DB 276 SUTG-----GVDRLRYERNKISAGYDGTGTFPGTWKS-----Y 308
QY 392 NFRFRSTNEQRTTLVA-----RGGLALNFRSIPQVDLIANA-RKGVRSYSHTVAT---- 441
DB 309 LNWNETENKRELVRSLKRDKWGLA-GQPRELKESNLILNSLLTPLGESHVTVGGEF 367
QY 442 -----ENLDFEYIGKSTPHPADGLSLIGGRLGHVYKIESGEGKTLH 483
DB 368 QSSSMKDGVVVLASTGETFPKQKXWSVFAEDWEHLTDALATAGSRYEHQFQGGH----- 422
QY 484 KASKTKFTGYAGAVVDLNDNNSI-----YLS--LSQLYTPQTNLDDADGKL-----L 527

DB 423 -----FSPRAYLVWDVADAWTLKGVTTGYKAPRMGQLHKGISGVGGKTNLLGNPDL 476
QY 528 KPRQNGQEVGYKSYMD--RLNARVSYFYRMKDKN-AAAPLNPNKKTRYAALGKRVME 584
DB 477 KPEESVSEAGV---YYDNPAGLNANVTGFTDFSNKIXVSYSINDNTNSVYNSGKARLH 533
QY 585 GVETISGAVTPKQW----IHAGYSYLSHQIKTASNRDDGIFLLMPKHSANLWTTYQVT 640
DB 534 GV--EFAGTL-PLMSEVDTLNLTWTRSEQRDGNK--GAPLSYTPHEHVMNAKLNWQIT 588
QY 641 PELTITGGV-----NAMSGITSSA---GMHAGGYATFDAAAYRFTPKLKL 683
DB 599 EVASWLGARYRGKTPRTQNYSSLSAVQKVYDEKGYLKAWTVVDAGLSKMTDALTL 648
QY 684 QINADNIFNRHY-----YARVGANT-FNIPGSERTWTANLRYSF 722
DB 649 NAAVNNLLNKDYSVLSYAGSAGSTLYAGDYFQTSSTTGYVIP--ERNYMSLNTQF 703
RESULT 5
US-09-668-113A-8
Sequence 8, Application US/09668113A
Patent No. 6410703
GENERAL INFORMATION:
APPLICANT: Russo, Thomas A.
TITLE OF INVENTION: Identification of A Vaccine Candidate from an
EXTRINSIC STRAIN OF E. COLI
FILE REFERENCE: 11520.0214
CURRENT APPLICATION NUMBER: US/09/668,113A
CURRENT FILING DATE: 2000-09-22
NUMBER OF SEQ ID NOS: 10
SEQ ID NO 8
LENGTH: 725
TYPE: PRT
ORGANISM: Escherichia coli
FEATURE:
US-09-668-113A-8

Query Match 7.2%; Score 273.5; DB 4; Length 725;
Best Local Similarity 19.8%; Pred. No. 1.7e-16;
Matches 167; Conservative 127; Mismatches 304; Indels 247; Gaps 40;
QY 5 MSVPRINMTAATVLAALSSSVFAAQTADLTVHLKQORSNAIVTEKNGDYSSFAVTVGT 64
DB 1 MRINKILWSLTVLLVGLNSQVSAKYS-----DNDDELTVVEA 39
QY 65 KIPASLRIPQSVSIITNQVKDR-NVDTPDLQARKTPGLRVLS-----DGRSSVY 116
DB 40 TAEQVLKQOP-GSVVITSEDIKTPPVNDLSDIIRKMFQVNLGNSASGTRGNRQIDIR 98
QY 117 ARGVEYSYINIDGLPAQMOS-----INGTLPNLFAPDRVEMRGSGLPDSS 162
DB 99 GMPENTLILIDGVPVTSRNSVYSWRGERDTRGDTNRWVPP-EQVERIEVIRGPAARVY 157
QY 163 SGEMGIYNLVRKPTKAFQGHAAAGFTHQYKAE-----DVSGSLNSDG--SV 211
DB 158 SGAAGVVNIITKRTNDW--HGSLSLYTNQPESEEGATRANFSLSGPLAGDALTTTL 215
QY 212 RGRV-----MAQTVGASPRPAEKNRHETFYAAADWDINPDTVLG--AGLYQQRH 260
DB 216 YGNLKTADSDWDINS PVGTKNAGHEGVRNKDINGVSWKLNPOQIILDFEVGYSRQG-- 273
QY 261 LAPVNGLPADANNKLSLPQHFV--GADWNKFKMNSHDVFADLKHFGNG-----GYGKV 314
DB 274 ---NIYAGDTQNSSSSAVTSLSAKSGKETNRLYQNYGI---TH---NGIWDWGQSRF 322
QY 315 GMRYSRDRDADSNYAFAGSKLGMKTPAGRPCNTADDKACAVGLGTETIKOKALAFDASYSR 374
DB 323 GYVYEKTN-----NTRMNEGLSGGEGEIRLAGEK-----FTTNRLS 358
QY 375 PFLGNTAN-----EFVIGADYNFRSTNEQRTTLVARGGALNFRSIPQVDLI 425
DB 359 SWRTSGELNIPLVNVDQTLTVGAENRDKLDDPSSTS-----LTVND-RDI----- 404


```

RESULT 9
PCT-US95-06994-9
? Sequence 9, Application PC/TUS9506994
? GENERAL INFORMATION:
? APPLICANT: Children's Hospital & Medical Center
? APPLICANT: University of Washington
? APPLICANT: Washington State University Research Foundation
? APPLICANT: TARR, PHILIP I
? APPLICANT: BILGE, SIMA S
? APPLICANT: BESSER, THOMAS E
? APPLICANT: VARY JR, JAMES C
? TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
? NUMBER OF SEQUENCES: 14

```

Qy	60	VTWGPKIPASLAREIPOSYSIITITNOQVANDANVPDOLAKRTPGRLVSLNDGSRSSVA--	11
Db	34	VVTIASSVEONIKMDPASTISVTOEDLOKRPVONLIKDLKEVPGVQLITNEGNRGVXXSI	93
Qy	118	RGVE--YSEVNDLPAOMOSINTELPN-----LFAADREVMKPSGSLPDS	163
Db	94	RGILSSSTLLIVDG--KRVNSRNAVFRHNPEDLWIMIXXPVDISIERIEVVRGPMSSLYGS	15
Qy	164	GEMGCIYNLVKRPPTKAFQGHAAAGFTHQYKAEDVSGSLNSDGSVGRGVAAQTVGAS	223
Db	152	DALGAVNIITKKIGCKK-----SGTVVDTITQ-----	180
Qy	224	PRPEKNNRHETFFAADWDINP--DTVLGA--GYLYOQRHLAPYNGLPADANKLPSL	276
Db	181	---EHRDRGDT-NGGFFTSGLPIDVLKMAKAYSLAKREKDDPQMSSTTDD--TGEPRI	234
Qy	279	PQHFVFGVADNM-KFKM--NSHDVFADLKHYFGNGGYGKVMRSYSDRDADS-----	325
Db	235	--EGFSSRDGNVEEAMTPNQNDHTA-----GT--GFDRQDSDSLDLKNNXXXX	283
Qy	326	---NYAFGSKLGKMTPAGEPGCNTADK-----ACA	354
Db	281	XXXXXXRLERQVSVSHN-----GMDYGTSELKYYGEKVENKNNXXXXXXXXXXXXXX	332
Qy	355	VELGTETKOKALAPDASYSRPFRLGNATNEFV-IGADY-----NRRSTYEOGRTT	404
Db	333	PGNSPFTSEBNTVDGKYYTLP--TAINQFLTVGGSGWRHDKLSDAVNLVTGTSSTKSAS	388

QY 405 LYARGGLALNEPRSPQVDLIANARKVGRYSHTVATENLDFEYIGKSG-----TTHP 457
DB 390 QVAL--FVEDEWRIPEPLALT--GVR-----MDHETYGHEWSPRAYLVNA 433
QY 458 ADGLSLIGGRGLGHVYKIESGEGKTLHKASKTKFTGYAGAVYDLNNSLYLSLSQLYTPQ 517
DB 434 TUVTVVKG-----WATAFAP-----SLLQLSPDW 459
QY 518 TWLADAGKL-----LKPRQNGQFVGG--YKGS-----YMDRLNARVSYF 555
DB 460 TNSCRGACKIVGSPDLKPTESWELGLYYMGEGWLGVESSVTVFRND-VKDRIGIS 518
QY 556 RMKDKNAAPLNPKTKRYAALGRV-----MEGVETEISGAVTPKQIHAAG 603
DB 519 RTSDEVN-AAPGQNFVGFETGANGRRIPVFSYNNVKNKARNOGVETELKIPFNDWKLSIN 577
QY 604 YSLHSQIKTASNRDDGIFLLMPKXSA-NLWTTVQVTPELTIGGVNAMSGITSSAGMH 662
DB 578 YTY--NDGRDVSNGENKPLSDI--PFLALEDWSFY-----VSGHYTCOKRADSATAKT 627
QY 663 AGGYATFDMAAAYRFTPKLKLQINADNI 690
DB 628 PGYTIWNTGAAGVQTKDKVKLRAGVLNL 655

RESULT 10

US-08-765-081-6
; Sequence 6, Application US/08765081
; Patent No. 5798260
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium
; OPERATING SYSTEM: MS-Windows 3.1
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,081
; FILING DATE: March 26, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: CHOR-1-10286
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 652 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Vibrio Cholerae
US-08-765-081-6

Query Match 5.6%; Score 212; DB 1; Length 652;

Best Local Similarity 19.6%; Pred. No. 7.4e-11;
Matches 147; Conservative 99; Mismatches 299; Indels 204; Gaps 32;
QY 57 SPAVTVGTGKIPAS-----LREIPQSVSIITNQKQNRNDVTF 93
DB 10 SLSVTTLGLMPSASAFQAQDATKTDVTVVTAAGYAQVQINAPASISVISREDLESRYRDV 69
QY 94 DQARKTPGLRLVLSNDGERS--SVYARGVYSEYINIDGLPAQMOSI-----NGTLP 142
DB 70 TDALKSPGVTVTGGDTTIDISIRMGSGNYTLILVDGKQTSRQTRPNSDGPGEIQGWL 129
QY 143 NLPAFDRVEMVGRSGSLFDDSGEMGGIVNLVRKPTKAFQGHAAAGCTHKKYKAEADVS 202
DB 130 PLOAIEKRIEVRGPMSTLYGSDALGGVINIITRKQQQWGSNVLSTVVQENRASGDEQS 189
QY 203 GSLNSDGSVRGRVMAQTVGASPRPAEKNNRH-----ETFYAADMDINDP--TVLGAG 253
DB 190 ANFFVTGPLSDALSILQVYVQTTQDEDEIEHGYGDKLSRLTSKLYQLNPDHQLQLEAG 249
QY 254 YLYQORH-----LAPYNGLPADANNKPLSLPQHVFGADWNKFKON-----SHDVPADLKH 304
DB 250 VSAQDRENNVGKSAQSGCGRCGTCN-----TD-NQYRRNHVAVSH-----288
QY 305 YFGNGYGVKVG-----MRYSDRDADS-----NYAFAGSKLGMKTPAGRPCNCTADDKAC 353
DB 289 ---QODWQGVGQSDTYLOYEENTNKSREMSIDNTVF---KSTLVAPIGEH-----332
QY 354 AVGLGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNFRSTNEQQTTLTYARGGLAL 413
DB 333 MLSFGVEGKHESLE-----DKTSNKI-----SSRTHISNTQW-----AGFIE 369
QY 414 NEFRSIPQVDLIANARKVGRYSHTVATENLDFEYIGKSTFHP-----ADGLSLIGG 466
DB 370 DEWALAEQFRLTFGGR-----LDHDKNYG-SHFSPRVGVWMLDPLWTVKG 414
QY 467 GRLGHYKIESG-ECKTLHKASKT--KFTGYAGAVY---DLNNDNSLYLSLSQLYTPQNL 520
DB 415 G-----VSTGFAPQLREVTPDWQVSG-GGNIYGNPDLPQFETSKINKELSLMTSTGSL 467
QY 521 DADGKLLKPRQNGQFVGYKGSYMDRLNARVSYFRMKDKNAAPLN-----PNN---KK 572
DB 468 AAS-----LTAFHNDFKDKITRVACPANICTAGPNQWGTATP 503
QY 573 TRYAALGRVMEGVETEISGAVTPKQIHAGYSYLHSQIKTASNSRDDGIFLLMPKHSAN 632
DB 504 TYRVNIDEAETYGAEATLSLPITESVELSSYTYTHSEQKSGNFAGRP--LLQLPKHLFN 561
QY 633 LWTTVQVTPELTIGGVN-----AMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQIN 686
DB 562 ANLSWQTTDLNLSWANLNRYKQEMQPEGGASNDDFIAPSYTFIDTGVYALTDTATIKAA 621
QY 687 ADNIFNRHY-YARVGGANTFNIPGSERTW 714
DB 622 VYNLFDQEVNVAEYG-----YVEDGRRVW 645

RESULT 11

US-09-098-082-6
; Sequence 6, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage

Db 576 VSGTPECTEDDAYYYRCS-----PYEKLDWQMKIDKARIRIGIELTGLANDKVA 627
Qy 592 GAVTPKQIHHAGYSYLSHQ-----IKTASNRDDGIFLLMPKHSANL 633
Db 628 SFPEBGMKLFGLSGVAKSKLSGNSLSLSTQPLKVIAGIDVESPEKMGVSRLLTYGAKK 687
Qy 634 WTTTYQVTPPELLTIGGVNAMSIGTSSAGMAGYATPDMAAAYRFTPKLKQIADNIFNR 693
Db 688 VKDAQTYVENKMGWGTPLQKQKVDYDPLNKSAY-VFDMYGFYKPAKLLTLRAGVYVLFNR 746
Qy 694 HY-----YARVG-----GANTFNIGSERBTWNLRYSF 722
Db 747 KTTTWSLRLGLYSSTTNNAVDRDCKGLDRYRAG--RNVAVSLEWKF 791
RESULT 14
US-08-326-670A-2
; Sequence 2, Application US/08326670A
; Patent No. 5698438
; GENERAL INFORMATION:
; APPLICANT: Scio, Inc., Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Hefitron, Fred
; APPLICANT: Nasait, Xavier
; TITLE OF INVENTION: A No. 5698438e1 Bacterial Hemoglobin Receptor Gene and Uses
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/326,670A
; FILING DATE: 18 OCT 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5698438nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 810-221-8317
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-326-670A-2
Query Match 4.9%; Score 185; DB 1; Length 792;
Best Local Similarity 19.6%; Pred. No. 3.3e-08;
Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;
Qy 14 AATVLAASSVFAAQTADLET-----VHIKQORSTNAIVTEKNGYSSFAVTVGT 64
Db 10 AATVGSIFGVPVPAADBAETTPVKAIVKAVKGGKRNPAVER-----VNLNR 60
Qy 65 KIPASLREIPQSYIITNOQVORNVDTFQDLARKTPGLRLVLSNDG-----RSSVYA 117
Db 61 IKQEMIRNDKIDVRYSTVDGLSDSGRHQKGFPAVRGVGNVGVSIDGVNLPDBEENSLYA 120
Qy 118 RGYEYSEYIDGIPAMQOSINGTLPLNLFADRYAVNRGPGSLFDDSGGEMGCIYVLRKR- 176

Db 121 R---YGNFNSRLSIDPELVR-----NIDYKADSFNTSGALGGVNIQTLOG 167
Qy 177 -----PTKAF-----OGHA-----AAGFGT-----HKQY 195
Db 168 RDLILPERQPGVMKMGKGYSTRNRWNTNLGFGVSNDNRVDAALLYSQRGHETESACKRGY 227
Qy 196 KAEADVSSLSNDGVSRRVMAQTVGASPRPAKNNRHETFPAAADWDINPDTVLGAGYL 255
Db 228 PVEGASGCA-NIRGSARG-----IPDPGQ--HKYHSFLGKLAIVQJNDNHRIGASLN 275
Qy 256 YQQRH-----LAPNGLPADANNK-----LPSLPQHYFVGA--DMNFKRN 294
Db 276 GQGGHYTYVEESYNYLASVYREADVDYNNRRNTNLFEYMTPESDRLSMVPAVDYQRTKYS 335
Qy 295 SHDVPADLKHYPNGGSGYKVMRYSRDDADSNYAFAGSKLGMKTPAGRCGCTADPKACA 354
Db 336 A-----VNYKGSFPIEDSSTLTNRVYNOKDDELTY-----NRSD----- 369
Qy 355 VGLGTEIKOKALAFDASYSRPRLAGNTANEFYIGADYNNFRSTNEGRTLLYARGLALN 414
Db 370 -----TRFKRITLRLD---SHPLQLGG-----GRRLSPKTPASRR 402
Qy 415 EPRSIPOVDLIANAR--KGVKGSHTVATENDDEGTYGKSTFHPADGLSLGGRLGHY 472
Db 403 DFEENLRDDYFSGRVVRRRTSSIQHPVKTNN--YGFSLSDQIQWNVDPSSRAGIRYDHT 459
Qy 473 KIESGE-GKTLHKASKYTK-----FTGYAGAVYDNL-----DNNSLYL--SLSQLY 514
Db 460 KMTPOELNMECHACDCTTPAANTYKMGSGFVGLAOLNQMVRGVYITTGIRYRNASERY 519
Qy 515 -----TPQTNLDAD-----GKLLK--PROGNOFEVGYKGYMDR--LNA 550
Db 520 FTYNHSGMWLPVPLKABRTTHTLSLOGRSEKGLDANLYQSNVRFNLFSEBOKLTTSG 579
Qy 551 RVS-----FYRKQKMAAFLPNNKTRRYAA--LGKVMESVE-----TEISGAT 595
Db 580 DVSCTQMNYYGMC-----NPYSEKLEWQMNIDKARIRIGIELTGLANDKVASFPV 632
Qy 596 PKQIHHAGYSYLSHQ-----IKTASNRDDGIFLLMPKHSANLMTY 637
Db 633 EGMKLGSLGYAKSKLSGNSLSLSTQPLKVIAGIDVESPEKMGVSRLLTYGAKKVKVA 692
Qy 638 QVTPPELLTIGGVNAMSIGTSSAGMAGYATPDMAAAYRFTPKLKQIADNIFNRHY-- 695
Db 693 QYTVENKMGWGTPLQKQKVDYDPLNKSAY-VFDMYGFYKPAKLLTLRAGVYVFNKRYTT 751
Qy 696 -----YARVG-----GANTFNIGSERBTWNLRYSF 722
Db 752 WSLRLGLYSSTNSVDRDCKGLDRYRAG--SRNVAVSLEWKF 792
RESULT 15
US-08-990-470A-2
; Sequence 2, Application US/08990470A
; Patent No. 6123942
; GENERAL INFORMATION:
; APPLICANT: Scio, Inc., Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Hefitron, Fred
; APPLICANT: Nasait, Xavier
; TITLE OF INVENTION: No. 6123942e1 Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehrnen Hubbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/990,470A
;; FILING DATE: 15-DEC-1997
;; CLASSIFICATION: 536
;; ATTORNEY/AGENT INFORMATION:
;; NAME: No. 6123942nan, Kevin E
;; REGISTRATION NUMBER: 35,303
;; REFERENCE/DOCKET NUMBER: 94,784-K
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 312-913-0001
;; TELEFAX: 312-913-0002
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 792 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-990-470A-2

Query Match 4.8%; Score 183; DB 3; Length 792;
Best Local Similarity 19.6%; Pred. No. 5.1e-08;
Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;

QY 14 AATVLAALSSSVFAAQTADLET-----VHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
DQ 10 AALVGSIFGNPVFAADEAATETTPVKAEVKAVRVKQGNAPAAVER-----VNLNR 60
QY 65 KIPASLREIPQSVSIITNOQVDRNVDTFDQLARKTGLRLVLSNDDG-----RSSVYA 117
DQ 61 IKQEMIRDNKDLVRYSTDVGLSDSGRHQKGFVAVRVEGNRGGVSDIGVNLDPSENSLYA 120
QY 118 RGVSEYSEYNIDGLPAQMQSINGTLPNLFAPDRVEMRGPGLFDSSGEMGIVNLVRKR- 176
DQ 121 R---YGNFNSRLSIDPELVR-----NIDIVKGADSFNTSGALGGGVNYQTLOG 167
QY 177 -----PTKAF-----QGHA-----AAGFGT-----HKQY 195
DQ 168 RDLLLPERQFGVMKNGYSTRNRENTLFGVSDNRDVAALLYSQRRGHETESAGKGY 227
QY 196 KAEADVSGSLNDSGVSGRVMAQTGASPRPAEKNNRHETFYAADWDINPDTVLGAGYL 255
DQ 228 PVEGAGSGA-NIRGSARG-----IPDPQ--HKYHSFLGKIAYQINDNHRIGASLN 275
QY 256 YQORH-----LAPYNGLPADANNK-----LPSLPQHVFGA--DMNKFOON 294
DQ 276 GQGHNYTVESYNLLASTWREADDVNNRRNTNLFYEWTPESDRLSWKVADVDYQTKVS 335
QY 295 SHDVFADLKHYPGNGGYGVKGYMRYSRDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACA 354
DQ 336 A---VNYKGSFPIDSSLTTRNYNQKLDLEY-----NRSMQ-----369
QY 355 VGLGTEIKOKALAFDASYSRPRLGTANTANEFVIGADYNFRSTNEQGRITLYARGGLALN 414
DQ 370 -----TRFKRITRLD---SHPLQLGG-----GHRLSFKTFASRR 402
QY 415 EFRSIPQVDLIANAR--KGVRGVSHVTATENLDEFGLYKSTFHPADGLSLIGGGRGLGHY 472
DQ 403 DENLNRRDYYFSGRVVRTSSIQHPVKTN---YGFSLSDQIQWINDVPSSRAGIRYDHT 459
QY 473 KTESGE-GKTLHKASKTK-----FTCYAGAVVDLN-----DNNSLYL--SLSQLY 514
DQ 460 KWTPELNAECHACDKTPPAANTYKWSGFGVGLAQLAQWVRGVYDITSGYRVPNASEVY 519
QY 515 -----TQTNLDAD-----GKLLK-PRQGNQFEVYKGSYMDR---LNA 550
DQ 520 FTYNHGSGNWLNPNLKABRTTHTTSLQGRSEKGTLDANLYQSYRNFSLSEQKLTTS 579
QY 551 RVS-----FYRMKDKNAAPLNPNKTRYAA--LGKRVMEGVE-----TEISGAVT 595
DQ 580 DVSCTQMYYNYGMS-----NPYSEKLEWQWQIDKARIGELTGLNLVDKVASFPV 632
```

```
QY 596 PKWQHAGYSYLHSQ-----IKTASNSRDDGIFLLMPKHSANLWTTY 637
DQ 633 BGWKLFGSLGYAKSLGCDNSLLSTQPLKVIAGIDYESPSEKMGVFSRLTYLGAKKVKDA 692
QY 638 QVTPELTIGGVNAMSGITSSAGMHAGGYATFDAMAAVRFPTKLQIQAQINAHNRYH-- 695
DQ 693 QYTVYENKMGWTPLOKKVKDYPWLNKSAI-VFDMYGFYKPVKNLTLRAGVNVNFRKTYT 751
QY 696 -----YARVG-----GANTFNIPGSERTWTANLRYSF 722
DQ 752 WDSLRLGLYSYSTTNSVDRDGKGLDRYRAP--SRNYAVSLEWKF 792
```

```
RESULT 16
US-08-817-707-2
; Sequence 2, Application US/08817707
; Patent No. 6277382
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 792 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-2
```

```
Query Match 4.8%; Score 183; DB 4; Length 792;
Best Local Similarity 19.6%; Pred. No. 5.1e-08;
Matches 173; Conservative 101; Mismatches 335; Indels 274; Gaps 39;

QY 14 AATVLAALSSSVFAAQTADLET-----VHIKQORSYNAIVTEKNGDYSSFAVTVGT 64
DQ 10 AALVGSIFGNPVFAADEAATETTPVKAEVKAVRVKQGNAPAAVER-----VNLNR 60
QY 65 KIPASLREIPQSVSIITNOQVDRNVDTFDQLARKTGLRLVLSNDDG-----RSSVYA 117
DQ 61 IKQEMIRDNKDLVRYSTDVGLSDSGRHQKGFVAVRVEGNRGGVSDIGVNLDPSENSLYA 120
QY 118 RGVSEYSEYNIDGLPAQMQSINGTLPNLFAPDRVEMRGPGLFDSSGEMGIVNLVRKR- 176
DQ 121 R---YGNFNSRLSIDPELVR-----NIDIVKGADSFNTSGALGGGVNYQTLOG 167
```



```

Db 587 THPKLLSANPEFGAQTGTWTASLAIRFKNLNLEIGWGRVYQKATGSILAAQKDRDG 646
Qy 663 -----AGVATFDAMAARFTPK--LKQINADNIFNRHYIYARVGGANTFNIPGSERTW 714
Db 647 KLENVVRQGFVNDVFNANKPLGKDTLNVNLSVNNVDFKFFY-----PHSQR-W 694
Qy 715 TANL 718
Db 695 TWTL 698

RESULT 18
PCT-US95-06994-7
; Sequence 7, Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IgA amino acid sequence,
; DESCRIPTION: wherein Xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:6.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: VIBRIO CHOLERAEE
PCT-US95-06994-7

Query Match 4.8%; Score 180.5; DB 5; Length 718;
Best Local Similarity 18.8%; Pred. No. 7.4e-08;
Matches 146; Conservative 92; Mismatches 285; Indels 255; Gaps 31;

Qy 57 SFATVGTGKIPAS-----LREIPQSVSLITNQVQKDRVD 91
Db 10 SLUSVTGLMFSASXXAFAQDATKTDETVMVTAAGYAQVIONAPASISVISREDLESRYR 69

```

```

Qy 92 TFDQARKTPGLRVLSDDDGRS-----SVYARGYSEYNIDGLPAQMOSI----- 137
Db 70 DVTDALKSVPQVTYXXXTGGDDTTDISIRGMSNYTLILVDGKRQTSRQTRPNSDGPGE 129
Qy 138 NGTLPNLPADPRVEMRGPGLPSSGEMGGIVNLVRKPTKAFQGH-----AA 187
Db 130 QGWLPPLOAERIEVIRGPMSTLYGSDAIGGVINIITRKQOQOQSGNVQLSTVVQENRAS 189
Qy 188 GFGTHQYKAEADVSGSLNSDGSVRGRVMAQTVCASPRPAE-----KNNRHETFYAAADW 242
Db 190 G-----DEQSANFFVTGPL--SDXALSLOVYQTTORDEDEIEHGYGDKSLASLT--SKLNY 242
Qy 243 DINPD--TVLGAGLYQQRH-----LAPYNGLPADANNKLPSPLOHVFVG-- 285
Db 243 QLNPDHQLQLEAGVSAQDRENNVGKSAQSGCXXXXXXRGTCSTNDQYRR--NHVAVSH 300
Qy 286 -ADNKKPMNSHDVPADLKHYFGNGGYKVMRYSDRDADS----- 325
Db 301 QGDWQ-----GVXGSDTYLYQEENTNKSREXXXXXXKXXXXXXX 343
Qy 326 -----NYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFADASYSRPRLG 379
Db 344 XXMSIDNTVP---KSTLVAPIGEH-----MLSGVEGKHESLE-----D 379
Qy 380 NTANEFVIGADYNRFRSTNEQRTTLYARGGLALNEFRSIPQVDLIANARKGVGYSHTV 439
Db 380 KTSNKI-----SSRTHISNTQW-----AGFIEDWALAEQPLRTFGGR----- 417
Qy 440 ATENLDEFGIYKSTFHP-----ADGLSLIGGRIGHYKIESGEGKTLHKASKTKFT- 491
Db 418 ----LDHDKNYG-SHFSRVYGVWNLDPMTVKG-----VSTG-----FRAPQLREVT 461
Qy 492 -----GVAGAVY---DLNDNNSLYLSOLYTPOTNLDADGKLLKPRGQNFVGYK 540
Db 462 PDWQVSGGGXXNIYGNPDLOPETSINKLSLWYSTGSLAAS----- 504
Qy 541 GSYMDDRNLNARVSFYRMKDKNAAAPLN-----PNN-----KKTRYAALGRKVMGVEVETRISG 592
Db 505 -----LTAFHNDFKDKITRVACPANICTAGPNQMGATPYRVNIDEAETYGAEATLSL 557
Qy 593 AVTPKQIHAGYSYLHSQIKTASNSRDDGIFLLMPKHSANLWTTYQVTPELTIGGGVNAM 652
Db 558 PITSEVELSSSYTYTHSEQSGNFAGRP--LLQLPKHLFNANLSWQTTDRLNSWANLNYR 615
Qy 653 SGITSSAGMEAG-----GYATPDMAAAYRFTPKLKLQINADNIFNR 693
Db 616 GKXXEMQPEGGASNDDFIAPXXXXXXSYTFIDTGTVTALTDTATIKAAVYNLFDQ 673

RESULT 19
US-08-537-361E-2
; Sequence 2, Application US/08537361E
; Patent No. 6121037
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```


OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E
FILING DATE: 02-OCT-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: No. 6121037nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-361E-2

Query Match 4.8%; Score 180; DB 3; Length 792;

Best Local Similarity 19.4%; Pred. No. 9.7e-08;
Matches 172; Conservative 104; Mismatches 328; Indels 284; Gaps 42;

QY 14 AATVLAALSSVFAAQTADLETVHI-----KQSRSYNAIVTEKNGDYSSFAVTGVT 64
DB 10 AALVGSIFGNPVPFAADEATETTPVKAAYKAVRKGQRNAPAAVER-----VNLNR 60
QY 65 KIPASLREIPQSVSIITNOQVDRNDVTFDQARKTGRLVLSNDG-----RSSVYA 117
DB 61 IKQEMIRDMKDLVRYSTVDGLSDSGRHQKFAVRGEGNRVGVSIQVNLDPSEENSLYA 120
QY 118 RGYEYSEVNIIDGLPAMQSIINGTLPLPAFDREYVWRGSPGLDSSGEMGIVNLVRRK- 176
DB 121 R---YGNFNSRLSIDPELVR-----NIDIVKADSFNTSGALGGGVNQTLOG 167
QY 177 -----PTKAF-----QGHAF-----AAGFGT-----HKQY 195
DB 168 RDLLEPERQFGVMKMGYSTTRNREWTNTLFGVSNDRVDAALLYSQRGHETESACKRGY 227
QY 196 KAEADVSGSLSDSGYRGVMAQTCVASPPRAKNNRHETFAAADMDINPDTVLGAGYL 255
DB 228 PVEGAGSGA-NIRGASRG-----IDPPSQ--HKYHSFLCKIAYQINDNHRIGASLN 275
QY 256 YQORH-----LAPYNGLPADANNK-----LPSLPQHVFYGA--DMNKFKRN 294
DB 276 GQGGHNYVVEESYNLLASTYWRADVDNRRRTNLFYEWTPESDRLSMVKADVDYQKTKVS 335
QY 295 SHDVFADLKHYFGNGGYGKVMRYSDRDADSNYAFAGSKLGMKTPAGRCQNTAADKACA 354
DB 336 A-----VNYGSGFPIEDSSTLTIRYNNOKDDELY-----NRSMD----- 369
QY 355 VGLGTEIKOKALAFDASYSRPFELGNTANEFVIGADYNFRSTNEQGRITLYARGGIALN 414
DB 370 -----TFPKRITRLD---SHPLOLQ-----GRHRLSKPTFSRR 402
QY 415 EFRSIPQVULLINAR--KEVRGSHTVATENLDEFGIYKSTHPADGLSLGGGLGHY 472
DB 403 DEENLRDYYRFSGRVVRTTSIOHPVKTTN--YGFSLSDQIOMNDVSSRAGIYYDHT 459
QY 473 KIESGE-GKTLHKASKTK-----FTGYAGAVYDLN-----DNNSLYL--SLSOXY 514
DB 460 KMTPEQLNMGACADCTIPRAANTYKMGSGFVGLAALNOAMRQYIITSGIRYPNASEYV 519
QY 515 -----TQTNLDAD-----GKLLK--PROGNOFEVGYKGSYMDR--LNA 550
DB 520 FTYNHSGMWLPPNPKAERTTHTLLOGRSEKGLDANLYOSNRYNRLSEBQKLTTSG 579
QY 551 RVS-----FYMKDKNAALPLNPNKKTIRYAA--LGKVMESVE-----TEIGAVT 595
DB 580 DVSCTQMNYYYGCS-----NPYSEKLEWONOIDKIRIGIELTGLRLNDVKVASFVP 632

QY 596 PKEQIHAGYSYHSQIKTASNRDGIFFLMP-----KHSANLW-----TTY---- 637
DB 633 EGMKLFSGSLGVAKSKL-----SGDNSILFTQPLKVIAGIDYSPSEKMGVSRFLYLGAK 687
QY 638 -----QVPELIIIGGVNAMSIGTSSAGNHAGGYATFDMAAYRFRPKLKQIOMDNIEN 692
DB 688 KYKDAQYTVYENKMGWGTPIQKKVKQYPMLNKSAV--VFDMYGFYKPKVKNLTLLRAGYVNVN 746
QY 693 RHY-----YARVG-----GANTFNIPEGSERTWNTANLRYSF 722
DB 747 RXYTWDLSRLGLYSYSTNYSVDRDGKGLDRYAP--SRNYAVSLERKF 792

RESULT 20

US-08-537-361E-6
Sequence 6, Application US/08537361E
Patent No. 6121037

GENERAL INFORMATION:
APPLICANT: Stojiljkovic, Igor

APPLICANT: So, Magdalene

APPLICANT: Hwa, Vivian

APPLICANT: Heffron, Fred

APPLICANT: Nassif, Xavier

TITLE OF INVENTION: No. 6121037e1 Bacterial Hemoglobin Receptor

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSER: McDonnell Boehnen Hulbert & Berghoff

STREET: 300 South Wacker Drive, 32nd Floor

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/537,361E

FILING DATE: 02-OCT-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: No. 6121037nan, Kevin E

REGISTRATION NUMBER: 35,303

REFERENCE/DOCKET NUMBER: 94,784-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 312-913-0001

TELEFAX: 312-913-0002

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 792 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-537-361E-6

Query Match 4.7%; Score 179; DB 3; Length 792;

Best Local Similarity 19.8%; Pred. No. 1.2e-07;
Matches 175; Conservative 96; Mismatches 338; Indels 274; Gaps 40;

QY 14 AATVLAALSSVFAAQTADLET-----VHIKQSRSYNAIVTEKNGDYSSFAVTGVT 64
DB 10 AALVGSIFGNPVPFAADEATETTPVKAAYKAVRKGQRNAPAAVER-----VNLNR 60
QY 65 KIPASLREIPQSVSIITNOQVDRNDVTFDQARKTGRLVLSNDG-----RSSVYA 117
DB 61 IKQEMIRDMKDLVRYSTVDGLSDSGRHQKFAVRGEGNRVGVSIQVNLDPSEENSLYA 120
QY 118 RGYEYSEVNIIDGLPAMQSIINGTLPLPAFDREYVWRGSPGLDSSGEMGIVNLVRRK- 176
DB 121 R---YGNFNSRLSIDPELVR-----NIDIVKADSFNTSGALGGGVNQTLOG 167


```
QY 177 -----PTKAF-----QGH-----AAGFGT-----HKQY 195
; : : : : :
Db 168 RDLLEPQFGVMKNGYSTNRRENTNTLGFVGSNDRDVDAALLYGQRGHETESACKRGY 227
; : : : : :
QY 196 KAEADVSGSLSDGSGVRGVMAQTGASPRPAEKNNRHETFFAAADWDINPOTVLGAGYL 255
; : : : : :
Db 228 PVEGAGSGA-NIRGSRG-----IDPQO--HKYHSFLGKTAYQINDNHRIGASLN 275
; : : : : :
QY 256 YOORH-----LAPYNGLPADANNK-----LPSLPQHFVFGA--DMNFKKN 294
; : : : : :
Db 276 GOOGHNYTVEESYNLLASVREADDVNRNRNTNLFYEMTPESDRLSMWKADVDYQKTKVS 335
; : : : : :
QY 295 SHDVADLKHFGNGYGVKVMHRSYDRDADSNYAFAGSKLGMKTPAGPCGNTADDKACA 354
; : : : : :
Db 336 A-----VNYKSGPPIEDSSITLTRYNQKOLDEY-----NRSMD-----369
; : : : : :
QY 355 VGLGTBEIKOKALAFADASYRPRFLNGTAN--EFVIGADYNRFRSTNEQRTTLYARGGLA 412
; : : : : :
Db 370 -----TRFKKITLRLD---SHPLQGGRRHLSEKTFASRRDFENLRD-----YYFSGRV 418
; : : : : :
QY 413 LNEFRSIPQVDLIANARKGVRSYHTVATENLDEFGYKSTFHPADGLSLIGGRLGHY 472
; : : : : :
Db 419 VRTTSSI-----QHPVKTTN---YGFSLSDQIQWNDVFSSRAGIRYDHT 459
; : : : : :
QY 473 KIESGE-GKTLHKASKTK-----FTGYAGAVYDLN-----DNNSLYL--SLSQLY 514
; : : : : :
Db 460 KMTPOELNAEACHADCKTPPAANTYKWSGSGVGLAAQLNQAWRVGYDITSGYRVPNASEVY 519
; : : : : :
QY 515 -----TPQTNLADP-----GKLLK--PRQGNQFEGVKGYSMDR--LNA 550
; : : : : :
Db 520 FTYNHSNGWLNPNLKAERSTHTLTLGRSEKGTLDANLYQSNVRNLFSEQKLTTS 579
; : : : : :
QY 551 RVS-----FYRMKKNAAAPLNPNKTKRYAA--LGRKVMGVE-----TEISGAVT 595
; : : : : :
Db 580 DVSCOTMNYVYGCS-----NPYSEKLEWQWONIDKARIRGIELTGRNLNVKVASFVP 632
; : : : : :
QY 596 PKWQIHAGYSYLSHQ-----IKTASNSRDDGIFLLMPKHSANLWTY 637
; : : : : :
Db 633 EGKWLFGSLGYAKSKLSGDSNLSLSTOPLKVIAGIDYESPEKSGVFSRLTYLGAKVKXDA 692
; : : : : :
QY 638 QVTPELTIGGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKLQLOINADNIFNRHY-- 695
; : : : : :
Db 693 QYTVYENKNGWTPLOKKVVDYFWLNSAY--VFDVMGYFKPVKNLTLRAGVYVFNKRYTT 751
; : : : : :
QY 696 -----YARVG-----GANTFNIPGSERTWTANLRSF 722
; : : : : :
Db 752 WDSLRLGYSYSTNSVDRDGKGLDRYRAP--SRNVAVSLEWKF 792
; : : : : :
```

```
RESULT 21
US-08-628-434-2
; Sequence 2, Application US/08628434
; Patent No. 6265567
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Beucher, Margaret
; TITLE OF INVENTION: Isolated FrpB Nucleic Acid Molecule and Vaccine
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,434
; FILING DATE: 05-APR-1996
```

```
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/418,964
; FILING DATE: 07-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-3-P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-628-434-2
```

```
Query Match 4.6%; Score 173; DB 4; Length 713;
Best Local Similarity 19.0%; Pred. No. 3.7e-07;
Matches 163; Conservative 107; Mismatches 277; Indels 310; Gaps 41;

QY 8 FRINWTAATVLAALSSSVFAAQTADLETVHIKQORSYNAIVTEKNGDYSSFAVTVGTIP 67
; : : : : :
Db 6 FRLSLLSLTLAAGFAHAENANVALDTVTYKGRQ-----GSKI- 45
; : : : : :
QY 68 ASLREIPOSVISIINQOVKDRNVDTFQOLARKTPLGLRVLSNDGRSSVYARGYSEYNI 127
; : : : : :
Db 46 -----RTNIVTLQKDESTATDMRELLKEEPSIDFGGNGTQSQFLTLMGQNSVDI 97
; : : : : :
QY 128 DGLPAQMSQINGTLPNLPAFDRVEMRGPGLFDSSGEMGGIVN--LVRKRPTRKAFQHA 185
; : : : : :
Db 98 -----KVDNAYSDSQILYHQGRF--IVDPALVKVVSQKAGSA 134
; : : : : :
QY 186 AAGFCTHKQYKAEADVSGSLSDGSRGVMAQTGASP--RPAEKNRHRHTFFAAADWD 243
; : : : : :
Db 135 SAGI-----GATNGAIIAKTVDAODLLKGLDKN-----WG 164
; : : : : :
QY 244 INPDTVLGAGLYQQRHLAPYNGLPADANNKLPSPHOFVFGADWN---KFKMNSHDVFA 300
; : : : : :
Db 165 VR-----LNSGF-----AGNNGASYGASVF--GKEGNPDGLFSYRNDE-- 201
; : : : : :
QY 301 DLKHYFG-----NGG-----YKGYVMRYSDRD-----322
; : : : : :
Db 202 --KDYEAGKGFNDNGKGTVPYSALDKRSYLAKIGTTFDGDHRIVLSHMKDOHGRGIRTV 259
; : : : : :
QY 323 -----ADSNYAFAGSKLGM--KTPAG-----RPGCNTAD 349
; : : : : :
Db 260 REEFVASEKNSRITIKQAPSYRETTSQNTNLAYTGKDLGFEVKLDANAYVLEKKRYSAD 319
; : : : : :
QY 350 DKACAVGLG-----TEIKOKALAFD-----ASYSRPFRLNGTANEFVIGADY-- 391
; : : : : :
Db 320 KDXNGYAGNVKGNHTRIAFRSMNFPNDFSLAEQTLKYGINYRHOEIKPOAFLNSPEI 379
; : : : : :
QY 392 -NRFRSTNEQRTTLYARGGLALNEFRSIPQVDLIANARKGVRSYHTVATENLDEFGY 450
; : : : : :
Db 380 KDKKATNEEKKKN-----RENEKIAKAYRLTNPTK-----TDTGAY 416
; : : : : :
QY 451 GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKPTG-----YA 494
; : : : : :
Db 417 IEA--IHEIDGFTLGGRLYDRFKVTKTHDGTSSSSSNPSEFVQPREHWSFSASHNYA 475
; : : : : :
QY 495 G---AVYDLANDNNSLYLSLSQLYTPQTNLADAGKLLKPRQGNQFEGVKGYSMDRLNAR 551
; : : : : :
Db 476 GRSPLDYALQTHGKRGIIIS-----IADG--TKAERARNTIEGF--NYNDGTFAAN 522
; : : : : :
QY 552 VSYRMKDKKNAAAPLNPNKTKRYAA-----LGRKVMGVEVETIS---GAVTPKQIHAG 603
; : : : : :
Db 523 GSYFRQTIKDALA--NPQNRHDSVAVREAVNAGYIKNHGYELGASVYRTGGLTAK---VG 576
; : : : : :
QY 604 YSYLHSQIKTASNSRDDGIFLLMPKHSANLWTYQVT-----PELTIGG---VNAMS 653
; : : : : :
```



```

Db 577 VS--HSKRFYDTHDK-LLSANPERGAOVGRWTASLAVRPNLEIGMRRTYQXAV 633
Qy 654 GITSAGMH-----AGYATPDMAAAYRTPK-LKQINADNIENRHYARVAG 701
Db 634 GSILAAGQDRDOKLENVVRQGFVNDVFANMKPLKQDITLVNLVNNVFDKRY----- 688
Qy 702 ANTFNIPGSERTWTANL 718
Db 689 -----PHSOR-WTNTL 698

RESULT 22
US-08-613-009A-18
; Sequence 18, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loommore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-18

Query Match: 4.6%; Score 173; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;
Qy 3 QPMVSRIN--MTAAVIALSSVFA--AQTADLETYHIKQSRYNALVTEKNDYSS 57
Db 2 QOQHFRLLIILCLISMTALPAVAENVQAOQKQDITGVAKKQKTRDNEVTG---- 57
Qy 58 FAVTVGTKIPALREIPQSVSIITNQOVDRNVDTFDQAKRTPGRLRYLSNDDGRSVYA 117
Db 58 -----LGKIVKTAADTLSEKQVID-----IRDLTRYDPGIAVVEQGRGASSGS 100
Qy 118 -RGYESEYN--IDGLPAQMOG-----INGTLP-----NLFAPDR--VEVMKGPFG 158
Db 101 IKGMDKNRSLVVDGL-AQIQSYTAQALGRTTASGSSAINEIEENKAVAFISGSSNS 159
Qy 159 LFDSSGEMGIVNLVKKRPTKAFQGHAAAG-FGTHKQYAEADVSGSLNSDGSVRGRVNA 217

```

```

Db 160 VERQSGALAGSV-----AFQTKTADVDVIGEGQWGIQSTAYSGKNRGILQSLALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADMDINPDYLGAGYLVQGHHLAY----- 264
Db 211 GRIGAEALLITGRHA-----GEIRAHAAGRG-VQSFNRLAVDDGSKAYATVEE 262
Qy 265 ---NGLPADANNKLPSPQHFVVGADW-----NKRPMN--SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKPDVVGEDKQTVGTROVTGNRFLADPLSESSWLFRRPG 318
Qy 299 -PADLKHIFNGGSGYKVMRYSDRDA-----DSNVAPAGSKLQMKTPAGE---P 343
Db 319 RFENKHHYIG-GILERTQOTFPTDRTDWTVPALTKAVFDANOKQAGSLRGNGKXAGNHKYG 377
Qy 344 GCNTADKACAVG--LGTEIKQKALAFDASYSRPFRLGTANEFVIGAYNFRSTNEG 401
Db 378 GLFTSGENNAPVGAEGYGVFDEYHTKSRYGLEYVYTNADFD--TWADYARLS----- 429
Qy 402 RTLLYARGGLAL-NFERSIPOVDLIANARKGVRSHTVATENLDEFG-----LYGKSTH 456
Db 430 ----YDRQIGLDNHQ---QTHCSAD---GSDKYCRPADKPFSSYKSDRYIGRS----- 476
Qy 457 PADGLSLIGGKRLGHYKIESGEGKTLHKASKYF-----TGYAGAVYDINDNN 504
Db 477 -----HKLQAAFKKSFDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLIYLSIQLYTPQTNLDADGKLLK--RGNOFVGVYKKSVMYDDRANAVSPYRMKDXXA 562
Db 515 YYYQSANRAYSSKTPPQNNKKTSPNGRKNPYWSI-----GRNVVTRQICLFGNNT 568
Qy 563 AAPLNP--NKKTRYAALGKRWEGVETISGAVTPKW-QIHAGYSY----LHS---QIK 612
Db 569 YDCTPRGINSKSYAAVADVNRIG-----RAMDVAGALRYDVRSHSDGSS 617
Qy 613 TASN--SRDDGIFILMPHSANLWTTQVTPBLTIGGVNAMSGITSSAGMHAGYATF 669
Db 618 TGHRTLSWNAIVL--KPADWLDLYR-----TSTGRLLPFA-- 654
Qy 670 DMAAAYRTPKXK-LQINADNIENR 693
Db 655 -EMYGWRSQDKIKAVKIDPKSFK 678

RESULT 23
US-08-778-570B-24
; Sequence 24, Application US/08778570B
; Patent No. 6437096
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loommore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 03-JAN-1997

```


CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-664
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-778-570B-24

Query Match 4.6%; Score 173; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;

QY 3 QPMSVFRIN---MTAATVLAALSSSVFA--AQADLETVHIKQRSYNAIVTEKNGDYSS 57
DB 2 QOQHLFRLNILCLSLMTALPAYAENVQAOQAEKQOLDTIOVKAKKQKTRRDNEVTG--- 57
QY 58 FAVTVGTKIPASLREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRLVLSNDDGRSSVYA 117
DB 58 -----LGLVKVTADTSLSKQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
QY 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAPDR---VEVMRGPSSG 158
DB 101 IRGMDKNRVLSTVDGL-AQIQSYTAQAALGGTRTAGSSGAINIEYENKAVEISKGSNS 159
QY 159 LFDSSGEMGGIVNLVRKRPKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
DB 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGIQSKTAYSGKNRGLTQSLALA 210
QY 218 QTVGASPRAEKKNRHETFYAADWDINDPTVLGAGLYQQRHLPY----- 264
DB 211 GRIGGAELIIRTRHA-----GEIRAHAAGRG-VQSFNRLAPVDPDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSLPQHVFVGADW-----NKFKNV---SHDV----- 298
DB 263 ECKNG-----GHEKCKANPKDVUGEDKQRTVTRDTYGTGNRFLADPLSYESRSWLPFGF 318
QY 299 -PADLKHYFGNGYGVGMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
DB 319 RFENKRHYIG-GILERTQOTFTDRMTVPAFLTKAVFDANQOAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFLGNTANEFIGADYNFRSTNEQG 401
DB 378 GLFTSGENNAPVGAEGYGTGVFYDETHKSRYLEVYVYTNADKD--TWADYARLS----- 429
QY 402 RTTLVARGGLAL-NEPRSIPOVDLIANARKVGRVGSHTVATENLDEFG---IYKSTFH 456
DB 430 ----YDRQIGILDNIFQ---QTHCSAD-----GSDKYCRPSADRFYSYKSDRVIYES--- 476
QY 457 PADGLSLIGGRGLHYKIBESGKTLHKASKTKF-----TGAGAVYDLNDNN 504
DB 477 -----HKLQAFAFKKSFTDAKIRHNSLVNLGYDFRGSNLRHQD 514
QY 505 SLVLSLSQLYTQTNLDADGKLLK--ROGNQPEVGYKGSYMDRLNARVSYRMRKDKNA 562
DB 515 VYQSANRAYSKTTPQNNKGKTSNPGNEKNPYWVSI-----GRGNVTRQICLFGNNT 568
QY 563 AAPLNP--NKKTRVAALCKRMVGEVETISGAVTPKW-OIHAGYSY-----LHS---QIK 612
DB 569 YTDCTPRSRNGKSYAAVRDNRVLG-----RWADVAGRLRYDRSTHSDGSGVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLMTTYQVTPLETITGGGVNAMSGITTSAGMHAGGYATF 669
DB 618 TGTHTRLSNAGIVL---KPADWLDLTYR-----TSTGPRLPSPFA-- 654
QY 670 DAMAAYRFTPKLK-LQINADNLFNR 693

DB 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678
RESULT 24
US-09-059-584-24
Sequence 24, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loosmore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/059,584
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-24

Query Match 4.6%; Score 173; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.5e-07;
Matches 172; Conservative 99; Mismatches 292; Indels 242; Gaps 43;

QY 3 QPMSVFRIN---MTAATVLAALSSSVFA--AQADLETVHIKQRSYNAIVTEKNGDYSS 57
DB 2 QOQHLFRLNILCLSLMTALPAYAENVQAOQAEKQOLDTIOVKAKKQKTRRDNEVTG--- 57
QY 58 FAVTVGTKIPASLREIPQSVSIITNOQVKDRNVDTFDQARKTPGLRLVLSNDDGRSSVYA 117
DB 58 -----LGLVKVTADTSLSKQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
QY 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAPDR---VEVMRGPSSG 158
DB 101 IRGMDKNRVLSTVDGL-AQIQSYTAQAALGGTRTAGSSGAINIEYENKAVEISKGSNS 159
QY 159 LFDSSGEMGGIVNLVRKRPKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVMA 217
DB 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGIQSKTAYSGKNRGLTQSLALA 210

Db 618 TGHRTLSWAGLVL---KPADWLDLTYR-----TSTGRPLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSFNK 678

RESULT 26
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-124A-2

Query Match 4.6%; Score 172; DB 2; Length 915;
Best Local Similarity 21.4%; Pred No. 6,9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

Qy 3 QFMSVFRIN---MTAATVLAALSSSYFA--AQTALETVHIKQRSYNAIVTEKNGDYSS 57
Db 2 QOQHLFRNLCLSLMTALPAVAENVQAGQEKQDITQVAKKQKTRRDNEVTG--- 57
Qy 58 FAVTVGTKTPASLRPEIPQSVSIITNQVYKDRNVDFDQLARTPTGLVLSNDDGSSVYA 117
Db 58 -----LGKLVKTADTSLSEQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
Qy 118 -RGVEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAPDR---VEVMRGPSPG 158
Db 101 IRGMDKNRSLVTDGL-AQIQSYTAQAALGGTRTAGSSGAINEIYENVKAVEISKGSNS 159

Qy 159 LFDSSGEMGGIWNLVKRKPTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVGRVRMA 217
Db 160 VEOGSGLAGSV-----AFQTKTADDVIGEGROWGIOQSKTAYSGKNRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNRHRHETFYAAADWDINPDVVLGAGVLYQOORHLAPY----- 264
Db 211 GRIGGAELLIRTORHA-----GETRAHEAARG-VQSNRLAPVDDGSKYAYFIVEE 262
Qy 265 ---NGLPADANNKPLSPQHVFVGADW-----NKFQKN---SHDV----- 298
Db 263 ECKNG---GHEKCKANPKDDVGVGDEKQTVSTRTDYTPGNRFLADPLSYESRSWLPGRPF 318
Qy 299 -PADLKHYFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAFLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
Qy 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPFRLGNTANEFVIGADYNRFRSTNEQG 401
Db 378 GLFTSGENNAVPVGAEGYGTGVFYDETHTKSRYGLEBYVTNADKD---TWADYARLS--- 429
Qy 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVRGYSHTVATENLDEFG---IYCKSTFH 456
Db 430 ---YDRQIGLDNHQ---QTHCSAD---GSDKYCRPSADKPFYYSKSDRVIGES--- 476
Qy 457 PADGLSLIGGRLGHYKIESGEGTKLHKASKTF-----TGAGAVVDLNDNN 504
Db 477 -----HKLQAAFKKSPDTAKIRHNLSVNLGYDRFGSNLRHQD 514
Qy 505 SLYLSLSQLYTPQTNLDADGKLLKP--ROGNQFVGVYKGSYMDRLNARVSYFMRKDKNA 562
Db 515 YYYOSANRAYSLKTPPONNGKKTSPNGREKNPYWVSI-----GRGNVVTROI CLFGNNT 568
Qy 563 AAPLNP--NKKTRYAALGKVMGVEGTEISGAVTPKW-QIHAGYSY---LHS---QIK 612
Db 569 YDCTPRSINGKSYAAVRNVRVG-----RWADVGAGLRVDYRSTHSDGGSVS 617
Qy 613 TASN---SRDDGIFLLMPKHSANLWTTVQVTPELTIGGVNAMSGITSSAGMHAGGATVF 669
Db 618 TGTHRTLSWAGLVL---KPADWLDLTYR-----TSTGRPLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSFNK 678

RESULT 27

US-08-478-435-96
; Sequence 96, Application US/08478435
; Patent No. 5923232
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

Db 58 -----LGLKLVKTADTLKSKEQVLD-----IRDLTRYDPGLAVVEQGRGASSGVS 100
Qy 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLPFDR---VEVMRGPSSG 158
Db 101 IRGMDKNRVSILTVDGL-AQIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGNS 159
Qy 159 LFDSSGEMGGIVNLVRKRTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVM 217
Db 160 VEQSGGALAGSV-----AFQTKTADDVIGEGRGWGIQSKTAYSGKRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADWDINDPTVLGAGLYQQRHLAPY----- 264
Db 211 GRIGGAELLITGRHA-----GEIRAHEAAGRG-VQSFNRLAPVDPDGSKYAYFIVEE 262
Qy 265 ---NGLPADANNKLPSPHOFVFGADW-----NKFKMN---SHDV----- 298
Db 263 ECKNG---GHEKCKANPKKDVVGEDKQTVSTRDYTGPNRFLADPLSYESRSWLPFGPF 318
Qy 299 -FADLKHYFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQQTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
Qy 344 GCNTADKACAVG--LGTEIKQKALAFDASYSRPRLGNTANEFVIGADYNNFRSTNEQG 401
Db 378 GLFTSGENNAPVGAEGYGVFDETHTKSRYLEVYVYNADKD--TWADYARLS----- 429
Qy 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVRGYSHTVATENLDEFG----IYKSTTFH 456
Db 430 ----YDRQGIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYSKSDRVIYGES--- 476
Qy 457 PADGLSLGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLDNNN 504
Db 477 -----HKLQRAFKKSPDTAKIRHNLVSNLGLYDFRFGSLRHQD 514
Qy 505 SLVLSLSQLYTPQTNLDADGKLLKP--ROGNOFEYGVKGSYMDRLNARVSFYRMKDKN 562
Db 515 YYQSANRAYSKTTPQNNKKTSPNGREKNPYWIS-----GRGNVVTQICLFGNNT 568
Qy 563 AAPLNP--NKTRYAALGRVMGVETEISGAVTPKW-QIHAGYSY----LHS---QIK 612
Db 569 YTDCTPRSINGSKYAAVRDNRVLG-----RWADVAGRLYDYRSTHSDGGSVS 617
Qy 613 TASN---SRDDGIFLLMPKHSANLTYQVTPELTIGGVNAMSGITSGAGHAGGYATF 669
Db 618 TGTHRTLSSNAGIVL---KPADWLDLTYR-----TSTGRLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWSGDKIKAVKIDPEKSFNK 678

RESULT 29
US-08-478-373-96
; Sequence 96, Application US/08478373.
; Patent No. 5922841
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,373
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-373-96

Query Match 4.6%; Score 172; DB 2; Length 915;

Best Local Similarity 21.4%; Pred. No. 6.9e-07;

Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

Qy 3 QFMSVFRIN---MTAATVLAALSSVFA--AQTDADLETVHIKQORSYNAIIVTEKNGDYSS 57
Db 2 QQQHLFRNLILCLSLMTALPAYAENVQAGQAEKQLDTIQVKAKKQKTRRDNEVTG--- 57
Qy 58 PAVTVGTPKIPASLREIPOSVSIITNQVKRNVDTFQIARKEPCLRVLSNDDGRSSVYA 117
Db 58 -----LGLKLVKTADTLKSKEQVLD-----IRDLTRYDPGLAVVEQGRGASSGVS 100
Qy 118 -RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLPFDR---VEVMRGPSSG 158
Db 101 IRGMDKNRVSILTVDGL-AQIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGNS 159
Qy 159 LFDSSGEMGGIVNLVRKRTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVM 217
Db 160 VEQSGGALAGSV-----AFQTKTADDVIGEGRGWGIQSKTAYSGKRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNNRHETFYAAADWDINDPTVLGAGLYQQRHLAPY----- 264
Db 211 GRIGGAELLITGRHA-----GEIRAHEAAGRG-VQSFNRLAPVDPDGSKYAYFIVEE 262
Qy 265 ---NGLPADANNKLPSPHOFVFGADW-----NKFKMN---SHDV----- 298
Db 263 ECKNG---GHEKCKANPKKDVVGEDKQTVSTRDYTGPNRFLADPLSYESRSWLPFGPF 318
Qy 299 -FADLKHYFGNGGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQQTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
Qy 344 GCNTADKACAVG--LGTEIKQKALAFDASYSRPRLGNTANEFVIGADYNNFRSTNEQG 401
Db 378 GLFTSGENNAPVGAEGYGVFDETHTKSRYLEVYVYNADKD--TWADYARLS----- 429
Qy 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKGVRGYSHTVATENLDEFG----IYKSTTFH 456
Db 430 ----YDRQGIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYSKSDRVIYGES--- 476


```

Qy 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKYF-----TGAAGAVYLDNDNN 504
Db 477 -----HKLQAAFKKSPDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLVLSLSQLYTPQTNLDADKLLKP--RQGNQFEVGYKGSYMDRLNARVSFYRMKDXXA 562
Db 515 YYYQSANRAVSLKTPPQNNGKKTSPNGREKNPYWSI-----GRNVVTRQICLFGNNT 568
Qy 563 AAPLNPN--NKTTRYAALGRVMEGVETEISGAVTPKM-QIHAGSY---LHS---QIK 612
Db 569 YVDCPTRSINGKSYAAVNDVRLG-----RMADVGLRVDYRSTHSDGGSYS 617
Qy 613 TASN---SRDDGIFILMPKHSANLWTTQVTPBELTIGGAVNAMSGITSSAGMHAGSYATF 669
Db 618 TGHRTLSNNAGIVL---KPADWLDLTYR-----TSTGFRLPSEFA-- 654
Qy 670 DAMAAVFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678

```

```

RESULT 30
US-08-474-671-96
; Sequence 96, Application US/08474671
; Patent No. 6008326
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murrin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,671
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:V9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 96:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids

```

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-474-671-96

```

```

Query Match 4.6%; Score 172; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

```

```

Qy 3 QMSVFRIN--MTAATVAALSSVFA--AQTADLETVIHGORSYNALIVTEKNDYSS 57
Db 2 QOQHLEFRANILICLSMTLPAVAENVQAGQAEKQDLTIQVAAKQKTRDNEVTC---- 57
Qy 58 FAVTVGTKIPASLREIPQSVSIITNQVKNVDVTFDQARKTPGLRLVLSNDGSSVYA 117
Db 58 -----LGKLVKATDILSKQVLD-----IRDLTRYDPGIAVVEQGRGASSGYS 100
Qy 118 -RGYESEYN--IDGLPAMQMS-----INGTLP-----NLFAFDR--VEVWRGPG 158
Db 101 IGMMDKNRVSLTYVDGL-AQIQSYTAQALGTRTAGSSGAINIEIYENKAVEISKGSNS 159
Qy 159 LFDSSGEMGIYNLVKRPETKAFQGHAAAG-FGTHKOYKAADVSGSLNSDGSVRGRVVA 217
Db 160 VEOGSGALAGSV-----AFQTKTADVDVGEGRQMGIQSKTAYSGNRGLTQSIALA 210
Qy 218 QTVGASPRPAEKNNRHEETVYAADWDINDPVLGAGYLYQORHLAPY----- 264
Db 211 GRIGAEALLIRGRLA-----GEIRAEHAARG-VQSPFRLAVVDGSKYAFIVEE 262
Qy 265 ---NGLPADANNKLPSPHVFVYGADW-----NKRKNM--SHDV----- 298
Db 263 ECKNG---GHECKCANPKPDVYGEDKQTVSTRTDYGNRFLADPLSYESSRWLFRPGF 318
Qy 299 -PADLHYFGNGGCGVKNRSGDRDA-----DSNVFAGSKLQMKTPAGR---P 343
Db 319 RRENKHYTG-GILEFTQOTFOTRDNTVPAFLTKAVFDNQOAGSLRNGKXAGHKG 377
Qy 344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFRLGNTANEFVIGADYNFRSTNEOG 401
Db 378 GLFTSGENNAAPYGAEBYGTGVFDETHTKSRGLGEVYYTNADKD--TWADYARLS----- 429
Qy 402 RTTLVARGLAL-NEFRSIPQVDLIANARKVGRGSHVYATENLDEFG---LYGKSTH 456
Db 430 ---YDRQIGILDNHFO---QTHCSAD---GSDKYCRPSADKPFYSKSDRYLYGSS-- 476
Qy 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKYF-----TGAAGAVYLDNDNN 504
Db 477 -----HKLQAAFKKSPDTAKIRHNLVNLGYDRFGSNLRHOD 514
Qy 505 SLVLSLSQLYTPQTNLDADKLLKP--RQGNQFEVGYKGSYMDRLNARVSFYRMKDXXA 562
Db 515 YYYQSANRAVSLKTPPQNNGKKTSPNGREKNPYWSI-----GRNVVTRQICLFGNNT 568
Qy 563 AAPLNPN--NKTTRYAALGRVMEGVETEISGAVTPKM-QIHAGSY---LHS---QIK 612
Db 569 YVDCPTRSINGKSYAAVNDVRLG-----RMADVGLRVDYRSTHSDGGSYS 617
Qy 613 TASN---SRDDGIFILMPKHSANLWTTQVTPBELTIGGAVNAMSGITSSAGMHAGSYATF 669
Db 618 TGHRTLSNNAGIVL---KPADWLDLTYR-----TSTGFRLPSEFA-- 654
Qy 670 DAMAAVFTPKLK-LOINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSPNK 678

```

```

RESULT 31
US-08-483-577A-96
; Sequence 96, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena
; APPLICANT: Harkness, Robin

```


APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESS: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,577A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-511
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-577A-96

Query Match 4.6%; Score 172; DB 3; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;
QY 3 QFMSVPRIN--MTAATVLAALSSVFA--AQTADLETVHIKQSRYSNAIVTEKNGDYSS 57
DB 2 QOCHLRLNLCLSMTALPAENVAQQAQKQLDTIQVRAKKQKTRDNEVTG---- 57
QY 58 FAVTVGTKIPASLREIPQSVSIITNQVDRNVDTFDQLARKTPGLRLVLSNDGGRSSVYA 117
DB 58 -----LGLVKVTADTSLKEQVLD-----IRDLTRYDPGLAVVEQGRGASSGYS 100
QY 118 -RGYEYSEVN--IDGLPAQMQS-----INGTLP-----NLFAPDR---VEVMRGPSPG 158
DB 101 IRGMDKNRVSLTVDDL-AQIQSYTAQAALGGTRTAGSSGAINEIYENNVKAVEISKGSNS 159
QY 159 LFDSSGEMGIVNLVRKRTPTKAFQGHAAAG-FGTHKQYKABADVSGLSLSDGSGVRGYMA 217
DB 160 VEQGSLAGSV-----AFQTKTADTVIGEGRWGIGQSKTAYSGKNGRLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAADMINDPDTVLGAGLYQQRHLPAY----- 264
DB 211 GRIGGAELLIRTGRHA-----GEIRAHAAGRG-VQSFNRLAPVDGSKYAVFIVEE 262
QY 265 ---NGLPADANNKPSLPQHVFGADW-----NKFPMN--SHDV----- 298

Db 263 ECKNG-----GHECKAMPKQDVGVGDKRQTVSTRDYTPGNRFLADPLSIESRSMLFRGPF 318
QY 299 -FADLKHYFGNGGKGYGMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAFLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRRLGNTANEFVIGADYNRFRSTNEQG 401
Db 378 GLFTSGENNAFVGAEGYGTGFYDETHTKSRYGLEBYVVTNADKD--TWADYARLS----- 429
QY 402 RTTLYARGGLAL-NEPERSIPOVDLIANARKGVRSYSHTVATENLDEFG-----LYCKSTFH 456
Db 430 ----YDROGIGLDNHQ---QTHCSAD---GSDKYCRPSADKPFYSYKSDRVYIGES--- 476
QY 457 PADGLSLITGGGRLGHYKIESGEGKTLHKASKTKF-----TCYAGAVVDLNDNN 504
Db 477 -----HKLQAAPKSPDTAKIRHNLSVNLGYDRFGSNLRHQD 514
QY 505 SLYLSLSOLYTPQNLNLDADGKLLKP--RQGNQFVGVYKGSYMDRLNARVSYFMRMKQNA 562
Db 515 YYYQSANRAYSLKTPPQNNKKTSPNGREKNPYWVSI-----GRGNVVTTRQICLFGNNT 568
QY 563 AAPLNP--NKKTRYAALGKRVMEGVETEISGAVTPKW-QIHAGVSY-----LHS---QIK 612
Db 569 YDCTPRINSKSYAAVRDNRVLG-----RWADVAGLRVYDRSTHSDGGSVS 617
QY 613 TASN---SRDDGIFLLMPKHSANLWTTVQVTPELTIGGGVNMAGSITSSAGMHAGGYATF 669
Db 618 TGTHRTLSWNAVGL---KPADWLDLTVR-----TSTGFRLPSPFA-- 654
QY 670 DAMAARYFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGDKIKAVKIDPEKSFNK 678
RESULT 32
US-08-897-438-96
Sequence 96, Application US/08897438
Patent No. 6262016
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,438
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/483,577
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994

PRIOR APPLICATION DATA: US 08/175,116
 APPLICATION NUMBER: US 08/175,116
 FILING DATE: 29-DEC-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/148,968
 FILING DATE: 08-NOV-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Stewart, Michael I
 REGISTRATION NUMBER: 24, 973
 REFERENCE/DOCKET NUMBER: 1038-720
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (416) 595-1155
 TELEFAX: (416) 595-1163
 INFORMATION FOR SEQ ID NO: 96:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 915 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

Query Match	4.6%;	Score 172;	DB 4;	Length 915;
Best Local Similarity	21.4%;	Pred. No. 6.9e-07;		
Matches 172;	Conservative 100;	Mismatches 291;	Indels 242;	Gaps 43

Qy	3	QPMFVFRNT---	MTAAATIALASSVFA---	AQTADLEFVNH	KGORVYNA	IVTEKNUGYSS	57	
Dp	2	QOOHLFRNLICLSLMTLPA	YAEVNOGAGOEKQLOLTI	OVKAKKQKTR	DRNEVTG---		57	
Qy	58	FAVTGTRKIPASLREIP	OSVSIITNOQVKNRNVDT	FDOLARKTPGL	EVLEINDDGRSSVA	117		
Dp	58	-----	LGKLVYATDLSKEQULD	-----	INDLFRYPDGI	AVNGGRASSGYS	100	
Qy	118	-RGVEYSYV---	INGLPAWOS-----	INGTLP-----	NLFADR---	VEVMRPGSG	158	
Dp	101	IRGMDKNRVSLTVDGL-	AQIOSYPAOALG	GTRTAGSSGAIN	IEYENVA	VAETSKSNS	159	
Qy	159	LFDSSEMGGIVNLVR	KRPTAPCGHAAAG-	FCHHKQYKAL	ADVSGSLNDG	STRGVMA	217	
Dp	160	VEGSSGALAGSV----	AFQTKTADVIG	GGROWGISO	KTAYSGKRG	LQSIATA	210	
Qy	218	QTVGASPPAPAKNNR	HEFTVYAAADMDIN	PDVLBAGVLYOOR	HLAPY-----		264	
Dp	211	GRIGAEKLLIRTRHA	-----	GERAHEA	AGRG-VQSEFN	LAPVDDOSKTA	YFIVEE	
Qy	265	---NGLPADANNK	PSLEPOHVFVGA	W-----	-----	NKFKNN--SHDV	-----	
Dp	263	ECKNG----	GHEKCKANPK	KOVVEDKQ	QVTSRDTY	GPFRFLADPL	SYESRSLFPRPGF	
Qy	299	-FADLKTRFNGNGY	GKVGMYSDRDA	-----	DSNVA	FAGSLGKMT	PAGR--P	
Dp	319	REFEKRHITYG-	GILEERTQO	FTDRDMTPA	FLTRAV	PDANO	KQAGSLRGNGK	YAGNHKYG
Qy	344	GCMNTADPGACV	G--LGEIIOQK	ALPDA	SYSRFR	GLANTANE	VIGADN	PRFSTMEOG
Dp	378	GLFTSGENNAP	VGAEYGVGYDD	ETHTKSR	GLELEYVT	NDKD--TWADY	ARLS-----	
Qy	402	RTTYAARGCLL-	NEFRSIP	POVDLIANA	RKVRGY	SHTVAT	ENDEFG--	
Dp	430	----YDNGO	GIDLHNFQ--QTH	SAD---GSD	KTCRPSAD	KPFSYK	SRVILYGBE--	
Qy	457	PADGLSLGG	RLGHYKIES	GEGKTL	HKASKTKP-			
Dp	477	-----	HLKLLQA	FKKSPD	TAIKRHNL	SVNLGYDR	FGSNLHOD	
Qy	505	SLYLSLQ	LYTPQNL	LDADG	KLKP--RQ	QNEFVY	GKSGYMD	RLAARVS
Dp	515	YYYSANRA	YELKTRP	POUNGK	KTSPNG	EKRPVW	SVSI-----	
Qy	563	AAPLNPN--	NKTTYAAL	GKRVMS	GEVE	IEIGAV	TPKM-OIHAG	YSV----
Dp	569	YTDCTPS	INGKSYAA	AVRDN	VLG-----	EMADG	ALRYD	ASTHSDGGS

Oy	613	TASN---	SRDDGILAMPKISANIMTYYQVTP	PELLTIGGVANMAMSGITSSAGMAGGATP	663
		::	::	::	
		::	::	::	
		::	::	::	
Db	618	TGTRHTLSMANGIVL---	KPADMDLTYR	-----	TSIGFRLPSFA-- 654
Oy	670	DAMAAAYRTPKTK-LQINADNI	IFPR	693	
		::	::	::	
		::	::	::	
		::	::	::	
Db	655	-EMYGMRSGDKIKAVKIDPEKSEFK	678		

RESULT 33
US-08-637

Sequence 96, Application US/08637654
Patent No. 6358727
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena M
APPLICANT: Harkness, Robin E
APPLICANT: Schryvers, Anthony B
APPLICANT: Chong, pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Murlidin, Andrew D
APPLICANT: Klein, Michel H
TITLE OF INVENTION: HAEMOPHILUS TRANSFERRIN RECEPTOR GENES
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:

```

STREET 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: MSG 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,654
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/CA94/00616
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-595
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-637-654-96

Query Match      4.6%; Score 172; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43

QY  QPMSEFRIN--MTAAATYALALSSVFA--AQTADLETVHIGGQRSYNAVITEKGDYS 57
   |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db  2 QQQHLFRNIITCLSMTLTPAYAEVVQAGQDEKQLDTIIQYAKKKOXRDRDEVG---- 57

QY  FAVTVGIKPASLBEPISVSIIITNQYKDRNVDTFDQLARRTPGLRYLISNDGRSSVYA 117
   |:::||::||::||::||::||::||::||::||::||::||::||::||::||
Db  58 -----LGKLVTADTLLSREYLID-----IRDLTRYDPGIADVVEQGSGSSGY 100

QY  RGEYYSYN--INGLPROMOS-----INGTLP-----NLFAPDR---VEVNRGPBG 158
   |||::||::||::||::||::||::||::||::||::||::||::||::||
Db  101 IRGMDKNKSVSLTVGGL-AAIQSYTAQAALLGGRTTAGSSGAINEIEYENKAVEISKGSNS 159

```


QY 159 LFDSSGEMGIVNLVRKPTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVRMA 217
Db 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGQIOKSTAYSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAAADNDINPDTVLGAGLYQQRHLAPY-----264
Db 211 GRIGGAELLIRTRHA-----GEIRAHAAGRG-VQSFNRLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSPQHVFGADW-----NKFKN--SHDV-----298
Db 263 ECKNG---GHECKKANPKKDVGVGEDKQTVSTVDTGPNRFLADPLSVESRSLWLRPGF 318
QY 299 -FADLKHYFGNGGYGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRLGNTANEFVIGADYNFRSTNEQG 401
Db 378 GLFTSGENNAFVGAEGYGVFDETHTKSRYLEVYVYNADKD--TWADYARLS-----429
QY 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKVGRGYSHTVATENLDEFG----IYKSTFFH 456
Db 430 ----YDRQIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYKSDRVIYGES---476
QY 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLNDNN 504
Db 477 -----HKLQAAPFKKSFDTAKIRHNSVNLGYDRFGSLRHQD 514
QY 505 SLVLSLSQLYTPQTNLDADKLLKP--RCNQFPEVGYKGSYMDRLNARVSPYRMDKNA 562
Db 515 YYYQSANRAYSLKTPQNNKKTSPNGREKNPYVSI-----GRGNVVTQICLFGNNT 568
QY 563 AAPLPN--NKTRVAALGKRMVGEVETISGAVPKM-QIHAGYSY-----LHS---QIK 612
Db 569 YDCTPRSNGSKSYAARNDRVLG-----RWADVAGRLYRSTHSDGSGVS 617
QY 613 TASN---SRDDGIFLMPKHSANLWTVQVTPELTIGGVNAMSGITSGMHAGGYATF 669
Db 618 TGTHRTLSNAGIVL---KPADWLDLTYR-----TSTGRLPSFA--654
QY 670 DAMAAYRTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSGDKTKAVKIDPEKSFNK 678

RESULT 34

US-08-649-518-96
; Sequence 96, Application US/08649518
; Patent No. 6361779
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Fele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 160
; CORRESPONDENCE ADDRESS:
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/649, 518
;; FILING DATE: 17-MAY-1996
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/337,483
;; FILING DATE: 08-NOV-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/483,577
;; FILING DATE: 07-JUN-1995
;; APPLICATION NUMBER: US 08/175,116
;; FILING DATE: 29-DEC-1993
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/148,968
;; FILING DATE: 08-NOV-1993
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Stewart, Michael I
;; REGISTRATION NUMBER: 24,973
;; REFERENCE/DOCKET NUMBER: 1038-608
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (416) 595-1155
;; TELEFAX: (416) 595-1163
;; INFORMATION FOR SEQ ID NO: 96:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 915 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; US-08-649-518-96

Query Match 4.6%; Score 172; DB 4; Length 915;
Best Local Similarity 21.4%; Pred. No. 6.9e-07;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;
QY 3 QFMSVPRIN---MTAATVLAALSSSVEA--AQTADLTGVHIKQORSYNAIVTEKNGDYSS 57
Db 2 QQQHLFRNLILCLSLMTALPAYAENVQAGQAEKOLDTIQVKAKKQKTRRDNEVTG----57
QY 58 FAVTVGTGKIPASLREIPQSVSIITNQVKDRNVDTFQQLARKTGLRLVLSNDDGRSSVYA 117
Db 58 -----LGKLVKTADTLSKEQVLD-----IRDLTREYDGIADVVEGRCASSGYS 100
QY 118 -RGYSEYN--IDGLPAQMS-----INGTLP-----NLFAPDR---VEVMRGPFG 158
Db 101 IRGMDKNRVSLTVDDL-AQIQSYTAQAALGGTRTAGSGAINETEYENVKAVEISKGSNS 159
QY 159 LFDSSGEMGIVNLVRKPTKAFQGHAAAG-FGTHKQYKAEADVSGSLNSDGSVRGVRMA 217
Db 160 VEOGSGALAGSV-----AFQTKTADDVIGEGRWGQIOKSTAYSGKNRGLTQSIALA 210
QY 218 QTVGASPRPAEKNNRHETFYAAADNDINPDTVLGAGLYQQRHLAPY-----264
Db 211 GRIGGAELLIRTRHA-----GEIRAHAAGRG-VQSFNRLAPVDDGSKYAYFIVEE 262
QY 265 ---NGLPADANNKLPSPQHVFGADW-----NKFKN--SHDV-----298
Db 263 ECKNG---GHECKKANPKKDVGVGEDKQTVSTVDTGPNRFLADPLSVESRSLWLRPGF 318
QY 299 -FADLKHYFGNGGYGKVMRYSDRDA-----DSNYAFAGSKLGMKTPAGR---P 343
Db 319 RFENKRHYIG-GILERTQOTFTDRMTVPAPLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377
QY 344 GCNTADDKACAVG--LGTEIKQKALAFDASYSRPRLGNTANEFVIGADYNFRSTNEQG 401
Db 378 GLFTSGENNAFVGAEGYGVFDETHTKSRYLEVYVYNADKD--TWADYARLS-----429
QY 402 RTTLYARGGLAL-NEFRSIPQVDLIANARKVGRGYSHTVATENLDEFG----IYKSTFFH 456
Db 430 ----YDRQIGLDNHFO---QTHCSAD---GSDKYCRPSADKPFYKSDRVIYGES---476
QY 457 PADGLSLIGGRLGHYKIESGEGKTLHKASKTKF-----TGAGAVYDLNDNN 504

Db 477 -----HKLQAAFFKSFDTAKTRHNLISVNLGYDRFGSNI RHOD 514
Qy 505 SLVLSLSQVLTPTQNDADGKLLK--ROGNOPEVGYKSYMDRLNARVSFYRMKDKN 562
Db 515 YYYQSANRAYSLSLTPPQNNKKTSPNGREKNPFWVSI-----GRGVNVTROCLGNN 568
Qy 563 AAPLNPN--NKTRVYALGKRVMEGVETEISGAVTPKW-QIHAGYSY---LHS---QIK 612
Db 569 YTDCTPRISNGKSYYAAVDNVLG-----RMADVAGLRDYRSTHSDGSVS 617
Qy 613 TASN---SPDDGFLIMPXHSANLMTTYQVTPPELLTIGGVMNMSGITSSAGMHAGYATP 669
Db 618 TGHRTLSMNAIVL---RPADWLDLTYR-----TSTGFRLLPSFA-- 654
Qy 670 DAMAAYRFTPKLK-LQINADNIFNR 693
Db 655 -EMYGWRSQDKIKAVKIDPEKSFNK 678

RESULT 35
US-08-363-124A-4
; Sequence 4, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Corneliussen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 909 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-363-124A-4

Query Match 4.5%; Score 171; DB 2; Length 909;
Best Local Similarity 18.2%; Pred. No. 8, 4e-07;
Matches 180; Conservative 128; Mismatches 327; Indels 356; Gaps 49;

Qy 3 QFMSVFRIMTAATVLAALSSVFA-----AQADLETVHIKGRSYNAIVTERKNGY 55
Db 2 QOQHLEFRNLILCLISMTAL--PYAENVOAEOAOEKOPLDTI QVAKAKQKTRDRNEVTG-- 57
Qy 56 SSFATVVGKIRIASLREIQSVSIITNOQVKBANDVTPQGLAKTGLVLSDDDRSSV 115
Db 58 -----LQKLVKSSDPTLSKEQVNL-----IRDLTRYDGIADVVEGGRASSG 98
Qy 116 YA-RGYEYSEYN--IDGLPAOMOS-----INGTLP-----NLPAFR--VEYWRGP 156
Db 99 YSIRGMDKRRVSLTVDGV-SQIQSYTAQAALGSTRAGSSGALNEIENYNAVEISKS 157
Qy 157 SGLFDSSGEMGGIVN----- 171
Db 158 NSSEYNGNLAGSVAFQRTAADIIIEGKOWIGSQKTAVSGKHALLTQSLALAGSGAE 217
Qy 172 --LVKRPYKAFOGHAAGFGTH-----KQYAEADV-----G 203
Db 218 ALLIYTKRRGRREIHAHKDAGKGVQSFNRLVLDEDDKKEGSGQSDISLCEBECHNGYAACKN 277
Qy 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNRH----- 233
Db 278 KLKEDASVDERKTVSTQYTGSNRLANPLEYGSQSLFPRGMHLDNRHYVGAIVLERTQ 337
Qy 234 ETF-----YAADWDINP-----DTVLGAGY---LY 256
Db 338 QTFDDTDMVPAVFTSE--DYVGSLSKGLKRGSDNKAERLFPVQEGESTIQIGYGVGPY 396
Qy 257 QQRHLAPYNGLPADANNKLPSLPQHYFVGADNMKFMNSHDVADLKHYPNGSGYKRVGM 316
Db 397 DERHTKRRYGV-----EYVYNAD-----KDTWAD---YARLSYRQGI 432
Qy 317 RYSDRPADSNVAFAGSKLGKMTKTPAGRGONTADDDKACAVGLGEITKOKALAPASRPF 376
Db 433 DLDNRLQOTHCHSDGSDKCKRPDGNKPYFSYSDRIT-----YESRNLFOAVKKAIF 485
Qy 377 RLGNTEANEFVIGADYNFRSTNEOGRTTLYARGGLANFRSIPQVLDIANNR---KGV 432
Db 486 DTAKIRHNLISNLGYDRFSS--QLSHSDYLLQ-----NAVQAYDLITPKKPPPNQS 535
Qy 433 RGYSHIVA---TENIDEGYIGKSTFHPADGLSLIGG-----RLGHYKIESGEG 479
Db 536 KNPFVYSIGKTTVNTSPPIPGFNNNTYDCTPRNIGNGNYAAVDNVLGRW-ADVGA 594
Qy 480 -----KTLKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLTPQT 518
Db 595 IRDYRSTSHSEDSKSVSTGTHRLNSMNAVGLKPTWMDLTYRSTGFRLLPSFEMYGWRA 654
Qy 519 NLDADGKLLKPRQNOFEVG--YKG-----SYMDRLNARVSF-YRMKDKN-----AA 563
Db 655 GBSLKTLDLKPEKSPFRREGIVFPGDFGMLNASYFNNAAYRDLIAFGYETRTQNGQTSAG 714
Qy 564 APINPNKKTRYAA---LGRVMEGVETEISGAVTPKW---QIHAGYSYLSHQIKTA 614
Db 715 DPGYRNGPKCTVAVAGINIGKIDMHGV-----WGGLPDGLYSTLAAYRIKVKDA 763
Qy 615 SNRSDGIGL-----LMPK-----HSANLM-----TTYO-----VTBELTI 645
Db 764 -DIRADRTVTSYLFDAVOPSRVYGLGYDHPDGMGINTMTFFYSVAKSVDELISQALL 822
Qy 646 GGGVNAMSGITSAGMHAGGYATFDAMAAYRFTPKLKLOINADNIFNRHY----- 695
Db 823 NGNANA---KKAASRRTFPWYTVDVSGYYNKKHLLTLAGAGVNNLNYVYVWENVRQTA 878
Qy 696 -----YARVGGANTFNIPGSERTWTANLRY 720
Db 879 GGAVNQHKNVGVNRYAADGRNVTFSLEMKF 909

RESULT 36
US-08-817-707-8
; Sequence 8, Application US/08817707
; Patent No. 6277382


```

;
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Naesif, Xavier
;
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
;
; TITLE OF INVENTION: Genes and Uses
;
; NUMBER OF SEQUENCES: 15
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 790 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-817-707-8

```

Query Match	4.5%;	Score 169.5;	DB 4;	Length 790;
Best Local Similarity	18.5%;	Pred. No. 9.2e-07;		
Matches 153;	Conservative 102;	Mismatches 336;	Indels 235;	Gaps 31;

QY	18	LAALSSSVF-----AAQPADLETHVHGKORSYNAIVTEKNGDYSFATVGTGTPKIPASLRE	72
DB	9	IAALVGSIFGNPVLAADEAATETTPVKAIEIKVVRVKQDLNAPATVERVNLGRIQOEMIRD	68
QY	73	IPQSVSIITNQOVKDRNVDTFDQLARKTPGLURVLNSDDG-----RSSVYARGVEYSEY	125
DB	69	NKDLVRYSTDVGLSDSGRHQRGFVARGVEGNRVGVSIDGVSLPDSEENSLYAR---YGNF	125
QY	126	NIDGLPAQMOSINGTLPNLPADFVRVWVRGPSGLFDSSGEMGGIVNLVRKPTKAPQGH-	184
DB	126	NSSLSDIDPELVR-----NIETAKGADSFNTGSCALGGVNY-----QTLLQGH	169
QY	185	-----AAAGFGTHKQYKABADVSGSLNS	207
DB	170	LLLDROFGVMMKNGYSSRNREWTNTLFGVSNDRVDAALLYSQRRGHETSAGERGYPV	229
QY	208	DGSRVGRVMAQTGASPRPAEKNNRHETFYAAADWDINDPTVLGAGLYLQQRH-----	260
DB	230	EGAGSGAIIRGSSRGIPDPSPK--HKYHNFLGKIAQINDKHIRIGSPFGNQOQHNYTIEES	287
QY	261	--LAPYAGLPADANNKLPSLPQHYPVFGADMKFKMNSHDVPADLKH-----FNGGG--	310
DB	288	YNLTASSNREADDVNR---RRNANLFEVWTPDSNWLSSLKADPDYQTTKVAAVNNKGSF	343
QY	311	---YKGVGRVSRDRDADSNYAFAGSKLGMTKTPAGRPGCNTADDKACAVLGTEIKQKALA	367
DB	344	PTDYSWTTRNYNQDLENLY-----NRSM-----TRFKRFTLR	377
QY	368	FDASYSRPFRLCNGTANEFVIGADYNRFRSTNEQGRITTLYARGGLALNEFRSIPQVDLIAN	427

```

Db      378 MD---SQPLQLG-----GRHRLSLKTFASRRREFENLRDDYYFS 413
Qy      428 AR--SGYRGYSHTVATENLDEFGIYGYKSTFHPADGLSLGGGRGLGHYKIESGE-GKTLHK 484
Db      414 ERVSRRTTSSIQHPVKTTN---YGFSLSDQIQWNVDFSSRADIRYDHTKWTTPQELNADCHA 470
Qy      485 ASKTK-----FTYAGAVYDLN-----DNNSLYL--SLSQYL-----TPQ 517
Db      471 CDKTPPAANTYKMGSGFVGLAAQLNQAWHVGVDYDITSYRVPNASEVYFTYNHSGNWLNF 530
Qy      518 TNLDA-----GKLLK-PROGNOPEVGYKGSVMDRLNARVS-----EYR 556
Db      531 PNLKAERSTHTLSLGRSEKGLDANLVQNNYR-NFLSEQLNLTSGDVCTQMNYYTG 589
Qy      557 MKDKNAAPLNPNKKTRYAA--LGRKRVMEGVE-----TEISGAVTPKQWIHAGYSYL 607
Db      590 MCS-----NPYSEKPEWQMONIDKARIRGLELTGRLNVTYKVASFVPGWKLFGLSLGYA 642
Qy      608 HSQ-----IKTASNRDDGGLFLMPKHSANLWTTYYQTVELTTGGGV 649
Db      643 KSKLSGDNLSLLTPPKVIAIGIDYESPSSEKMGVFSRLTYLGAKAKDAQYTVYENKGRGT 702
Qy      650 NMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQINADNIENRHY 695
Db      703 PLQKKVDYFWLKNKSAI-VFDMYGFYKLAKNLTLRAGVNVFNKRY 747

RESULT 37
US-08-613-009A-16
; Sequence 16, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-613-009A-16

```



```
Qy 234 ETF-----YAADWDINP-----DTVLGAGY-----LY 256
Db 338 QTFTDRMTVPAYFTSE--DYVPGSLKGLGKYSQDNKAERLFGQEGSTLQGIYGTGVFY 396
Qy 257 QORHLAPYNGLPADANNKLPSPHOFVGVGADWNKFMNSHDVPADLKHYPFGNGGKVGVM 316
Db 397 DERHTKNRYGV-----EYVHNAD-----KDTWAD-----YARLSYDRQGI 432
Qy 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
Db 433 DLDNRLQOCHSDGSKNCRPDGNKPYFYSKDRMI-----YEESRNLFOAVFKKAF 485
Qy 377 RLGNLTANEFVIGADYNRFRSTNEOGRRTLYARGGLALNEFRSIPQVDLIANARKVGRVYS 436
Db 486 DTAKIRHNLNLSINLGYDRFKS--QLSHSDYLLQNAVQAVDLYTPPKPPPPNGSKNPNYRVS 543
Qy 437 HTVATENLDEFGYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482
Db 544 IGTVTNTSPICRFGNNTYDCTPRNIGNGYAAVQDNVRLGRW-ADVGAIRYDYRST 602
Qy 483 HKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPOTNLDADGKL 526
Db 603 HSEDKSVSTGTHRNLSWAGVVLKPTWMDLYRSTGFRLPSPFAEMYGWRAGESLKTLD 662
Qy 527 LKPRQNGQEVG--YKG-----SYMDDLRLNARVSF-YRMKDKN-----AAAPLNPNKK 571
Db 663 LKPEKSFNREAGIVFGDFGNLEASYFNFNAYRDLIAFYETRTONGOTSASGDPGYRNAQ 722
Qy 572 KTRYAA--LGRVMGEGVETISGAVTPKW-----QIHAGYSLHSOIKTASNRDDGIF 623
Db 723 NARIAGINILGKIDHGV-----WGLPLDGLYSTLAYNRIKIKADA-DIRADRTF 770
Qy 624 L-----LMPK-----HSANLW-----TTYQ-----VTPELTIGGGYNAMSG 654
Db 771 VTSYLFDAVQPSRYVLGLYDHPDGTGWTGNTWFTYSKAKSVDELLGSQALLNGNANA--- 827
Qy 655 ITSSAGMHAGGATPDMAAAYRFTPKLKLOINADNIFNRHY-----YAR 698
Db 828 -KKAASRRTRPWPVYTDVSGYVNIKKHLTLRAGVYNLLNRYVYVTWENVROTAGGAVNOHKN 886
Qy 699 VGGANTFNIPGSERTWTANLRY 720
Db 887 VGVNRYAAPGRNYTFSLEMKF 908
```

RESULT 39

US-09-059-584-22

; Sequence 22, Application US/09059584

; Patent No. 6440701

; GENERAL INFORMATION:

; APPLICANT: Myers, Lisa E

; APPLICANT: Schryvers, Anthony B

; APPLICANT: Harkness, Robin E

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Du, Run-Pan

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/059,584

```
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-059-584-22
```

```
Query Match 4.5%; Score 169.5; DB 4; Length 908;
Best Local Similarity 18.0%; Pred. No. 1.2e-06;
Matches 177; Conservative 125; Mismatches 341; Indels 339; Gaps 46;
```

```
Qy 3 QMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQORSYNAIVTEKNGDY 55
Db 2 QOQHLFRNLCLSLMTAL--PVYAENVQAQAEKQLDTTQVKAKKQKTRDRNEVTG-- 57
Qy 56 SSFATVGTCTKIPASIREIPQSVSLITNQOVKDRNVDPDQLARKTPGLRVLNSDDGRSSV 115
Db 58 -----LGLVKSSDLSKEQVLN-----IRDUTRDPGIAVVEQGRGASSG 98
Qy 116 YA-RGEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAFDR---VEVMRGP 156
Db 99 YSRGMDKNRVSLLTVDGV-SQIQSYTAQAALGGTRTAGSSCAINEIEVENYKAVEISKGS 157
Qy 157 SGLFDSGEMGGIVN-----SGLFDSGEMGGIVN-----SGLFDSGEMGGIVN 171
Db 158 NSSEYGNAGLAGSVAFTQKTAADIIGEGKQWIOSKTAYSGKDHALQTSLAGRSGAE 217
Qy 172 ---LVRKPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
Db 218 ALLIYTKRGRREIHAHKDAGKGVQSFNRLVLEDDKXGGSQYRYFIVEECHNGYAACKN 277
Qy 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
Db 278 KLKEDASVKDERKTVSTQDYTGSNRLLANPLEYGSQSQWLPFGWHLDNRHYVGVALERTQ 337
Qy 234 ETF-----YAADWDINP-----DTVLGAGY-----LY 256
Db 338 QTFTDRMTVPAYFTSE--DYVPGSLKGLGKYSQDNKAERLFGQEGSTLQGIYGTGVFY 396
Qy 257 QORHLAPYNGLPADANNKLPSPHOFVGVGADWNKFMNSHDVPADLKHYPFGNGGKVGVM 316
Db 397 DERHTKNRYGV-----EYVHNAD-----KDTWAD-----YARLSYDRQGI 432
Qy 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
Db 433 DLDNRLQOCHSDGSKNCRPDGNKPYFYSKDRMI-----YEESRNLFOAVFKKAF 485
Qy 377 RLGNLTANEFVIGADYNRFRSTNEOGRRTLYARGGLALNEFRSIPQVDLIANARKVGRVYS 436
Db 486 DTAKIRHNLNLSINLGYDRFKS--QLSHSDYLLQNAVQAVDLYTPPKPPPPNGSKNPNYRVS 543
Qy 437 HTVATENLDEFGYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482
Db 544 IGTVTNTSPICRFGNNTYDCTPRNIGNGYAAVQDNVRLGRW-ADVGAIRYDYRST 602
Qy 483 HKASKTKFTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPOTNLDADGKL 526
Db 603 HSEDKSVSTGTHRNLSWAGVVLKPTWMDLYRSTGFRLPSPFAEMYGWRAGESLKTLD 662
Qy 527 LKPRQNGQEVG--YKG-----SYMDDLRLNARVSF-YRMKDKN-----AAAPLNPNKK 571
Db 663 LKPEKSFNREAGIVFGDFGNLEASYFNFNAYRDLIAFYETRTONGOTSASGDPGYRNAQ 722
Qy 572 KTRYAA--LGRVMGEGVETISGAVTPKW-----QIHAGYSLHSOIKTASNRDDGIF 623
Db 723 NARIAGINILGKIDHGV-----WGLPLDGLYSTLAYNRIKIKADA-DIRADRTF 770
Qy 624 L-----LMPK-----HSANLW-----TTYQ-----VTPELTIGGGYNAMSG 654
Db 771 VTSYLFDAVQPSRYVLGLYDHPDGTGWTGNTWFTYSKAKSVDELLGSQALLNGNANA--- 827
Qy 655 ITSSAGMHAGGATPDMAAAYRFTPKLKLOINADNIFNRHY-----YAR 698
Db 828 -KKAASRRTRPWPVYTDVSGYVNIKKHLTLRAGVYNLLNRYVYVTWENVROTAGGAVNOHKN 886
Qy 699 VGGANTFNIPGSERTWTANLRY 720
Db 887 VGVNRYAAPGRNYTFSLEMKF 908
```


QY 527 LKPRQNOFEVG--YKG-----SYMDRLNARVSF-YRMKDN-----AAPLIPNKK 571
 Db 663 LKPEKFNREBAGIVFKGDFGNLEASYNNAAYRDLIAFGYETRTQNOQTSAGDPGYRMAQ 722
 QY 572 KTRVAA---LGKRVMEGVETEISGAVTPKW-----QIHAGSYLHSQIKTASRPDGI 623
 Db 723 NARIAGINILGKIDMHGV-----WGGIPDGLYSTLAINRKVKDA-DIRADRTF 770
 QY 624 L-----LMPK-----HSANLW-----TTYQ-----VTPELTIGGVNAMS 654
 Db 771 VMSLFDVAVQPSRYVLGLGDHPDGIWGINMTFTYSKAKSVDELISQALINGNANA--- 827
 QY 655 ITSSAGMHAGGYATFDMAAAYRTPKLQIQINADNIFNRHY-----YAR 658
 Db 828 -KKAASRRTRPWWYTDVSGYNIKKHLLRAGVYNLLNRYVTWENVROTAGGAVNQHK 886
 QY 699 VGGANTFNIPGSERTWTANLRY 720
 Db 887 VGVYNRVYAAPGRNRYTFSLEMKF 908

RESULT 40
 US-08-487-890A-94
 ; Sequence 94, Application US/08487890A
 ; Patent No. 5708149
 ; GENERAL INFORMATION:
 ; APPLICANT: Loomore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Mardin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: 6th Floor, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/487,890A
 ; FILING DATE: 07-JUN-1993
 ; CLASSIFICATION: 435
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEPHONE: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 94:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 908 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-487-890A-94

Query Match 4.4%; Score 167.5; DB 1; Length 908;
 Best Local Similarity 18.2%; Pred. No. 1.8e-06;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QPMSVRIMTATVLAALSSVFA-----AQTDLETVHIKGRSINATYENKGOY 55
 Db 2 QOQHFLRNLILCLSLMTAL--PYAENVQAEQAEQKQDLTIQVAKQKQTRRDNETG-- 57
 QY 56 SSPAATVGTIKIPASLREIPQSVSIITNQCKDRNVDPOLARKTGLRVLNSDDRSSV 115
 Db 58 -----LKLKVSDDTSLKEQYL-----INDLIRYDGIATVEGGRASSG 98
 QY 116 VA-RGYEYSEYN-IDGLPAQMS-----INGTLP-----NLPAFDR---VEVMKRP 156
 Db 99 YSIRGMDKRVSLTVDGV-SQIQSYTAQAALGTRTAGSSGAINEIEYENVKAVEISK 157
 QY 157 SGLFSSGEMGIVN----- 171
 Db 158 NSSEYNGALAGSVAFQYTTAADIIGEKQWGIQSKTAYSGKDHALTQSLALAGRS 217
 QY 172 ---LVKRPTKAFQGHAAAGFGTH-----KQYK-----AEADVSG 203
 Db 218 ALLITYKRGREIHAHKDAGKGVQSFNRLVDEBKKEGSOYRIFYVEECHNGVYAC 277
 QY 204 SLNSDGSV--RGRVMAQTVGASPR---PAE-----KNRH----- 233
 Db 278 KLEKEDASVDERKTQVSTQYVTSNRLANPLEYSGQSWLFRGMHLDNHVYGAVALERT 337
 QY 234 EFT-----YAAADWQINP-----DYVLGAGY---LY 256
 Db 338 QFTEDTDMTPVPAFTSE-DYVGSLSKLGKYSQDNKAERLFPQGEESTIQGIGYGTVEY 396
 QY 257 QORHLAPYNGLPADANNKLPSPLOHVFVAGADNMKFRGNSHDVPADLKHFVNGGYK 316
 Db 397 DEHHTNRYGV-----EYVHNAD-----KDTWAD---YARLSYDRQGI 432
 QY 317 RYSDRPADSNYAFAGSKLGMKTPAGRPGCNTADKACAVGLTEIKOKALAPDAS 376
 Db 433 DLNRLQOQTHCHSDGDKNCRPDGKPKYFYSKDRNI-----YESNLLQAVPKAF 485
 QY 377 RLGNANAEVIGADNRPFRSTNEQRTTLYARGGLALNEFRSIPQVDLIANR---KV 432
 Db 486 DPAKRHNLSINLGYDRFES--QLSHSDYLLQ-----NAVQAYDLITPKKPPFPNGS 535
 QY 433 RGVSHTV---TENLDERGIYKSGTFHPADGLSLIGG-----RLGHYKIESG 479
 Db 536 KNPYRVSIGKTTVNTSPICRFGNNTYTDCTPRNIGNGVYAAVQDNVRLGRW-ADVG 594
 QY 480 -----KTLHAKSKYKFTG-----YAGAV-----YDLNUNSLYL---SLQLYTPOT 518
 Db 595 IRYDYSRTHSEBKSYSTGHRNLSMNAAGVVLKPFMTMDLTYASTGFRLPSPFAEMYG 654
 QY 519 NLDADGKLKPRQNOFEVG--YKG-----SYMDRLNARVSF-YRMKDN-----AA 563
 Db 655 GSELKTLDLKPKESFUREAGIVFKGDFGNLEASYNNAAYRDLIAFGYETRTQNOQTSAG 714
 QY 564 APLNPNKKTRVAA---LGKRVMEGVETEISGAVTPKW-----QIHAGSYLHSQIKTAS 615
 Db 715 DPGYRNAQNAIAGINILGKIDMHGV-----WGLPDGLYSTLAINRKVKDA- 762
 QY 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 Db 763 DIRADRTFTSYLFDVAVQPSRYVLGLGYHPGICINMTFTYSKAKSVDELISQALIN 822
 QY 647 GGVNAMSIGITSSAGMHAGGYATFDMAAAYRTPKLQIQINADNIFNRHY----- 695
 Db 823 GNANA---KKAASRRTRPWWYTDVSGYNIKKHLLRAGVYNLLNRYVTWENVROTAG 878
 QY 696 -----YARVGNANTFNIPGSERTWTANLRY 720
 Db 879 GAVNQHKVGVYNRVYAAPGRNRYTFSLEMKF 908

RESULT 41
 US-08-478-435-94
 ; Sequence 94, Application US/08478435
 ; Patent No. 5923233
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/478,435
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/337,483
 ; FILING DATE: 08-NOV-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/175,116
 ; FILING DATE: 29-DEC-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/148,968
 ; FILING DATE: 08-NOV-1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Stewart, Michael I
 ; REGISTRATION NUMBER: 24,973
 ; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (416) 595-1155
 ; TELEFAX: (416) 595-1163
 ; INFORMATION FOR SEQ ID NO: 94:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 908 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-478-435-94

Query Match 4.4%; Score 167.5; DB 2; Length 908;
 Best Local Similarity 18.2%; Pred. No. 1.8e-06;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;
 QY 3 QPMSVFRINWTAATVLAALSSVFA-----AQTADLETVHKQORSYNAIVTEKNGDY 55
 DB 2 QOCHLFRNLCLSLMTAL--PYVAENVOAQOEKQOLDTIQVAKKQKTRRDNEVTG-- 57
 QY 56 SSFAVTGKIPASLREIFQSVSIINQKVRNDVTFDLARKTPGLRLVLSNDDGRSSV 115
 DB 58 -----LGLKVSDDTSKEQVLN-----IRDLTRYDPGIAVVEQGRGSSG 98
 QY 116 YA-RGYEYSEYN--IDGLPAQMOS-----INGTLP-----NLFAFDR-----VEVMRGP 156
 DB 99 YSIRGMDKNRSLTVDGV--SQIQSYTAQAALGGTRTAGSSGAINIEYENKVAIVEISKGS 157
 QY 157 SGLFSSGEMGGINV----- 171

Db 158 NSSEYNGALAGSVAFTQTAAADIIGEGKQWGIOSKTAYSCKOHALTQSLALAGSSGAE 217
 QY 172 ---LVKRKPTKAFQGHAAAGFGTH-----KQYK-----AEDVSG 203
 Db 218 ALLIYTKERGRIHAHKDAGKGVQSPNRLVLDEKKEGSGYRYFIVEEECHNGYAAACKN 277
 QY 204 SLNSDGSV---RGRVMAQTVGASPR---PAE-----KONRH----- 233
 Db 278 KLKEDASVKDERKTVSTQDYTGSNRLLANPLEYGSQSWLFRPGWHLNDRHYVGVAVLRTQ 337
 QY 234 ETF-----YAAADWDINP-----DTVLGAGY-----LY 256
 Db 338 QTFDTRMTVPAYFTSE--DYVPGSLKGLGKYSQGNKAEERLFVQEGEGSTLOGIGTGVFY 396
 QY 257 QQRHLAPYNGLPADANNKLPSPQHFVFGADWNKFQKNSHDVFADLKHVFGNGGYGVKGM 316
 Db 397 DERHTKNRYGV-----EYVYHNAD-----KDTWAD-----YARLSYDRQGI 432
 QY 317 RYSDRADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTETIKKALAFDASYSRPF 376
 Db 433 DLNRLQOQTHCSHDSQKNCRPGDKNPKPYFYKSDRMI-----YBSRNLFOAVFKKAF 485
 QY 377 RLGNANTANEFVIGADYNRFRSTNEQGRITLYARGGGLALNEFESI PQVDLIANAR-----KGV 432
 Db 486 DTAKIRHNLNLSINLGYDRPKS--QLSHSDYLYQ-----NAVQAYDLITPKKPPFPNGS 535
 QY 433 RGYSHTV-----TENLDEFGIYKSTFHPADGLSLIGGG-----RLGHYKIESGEG 479
 Db 536 KDNPRVSIKGTVTNTSPICRFGNNTYDCTPRNIGGNGYAAVQDNRVLRGW-ADVGAG 594
 QY 480 -----KTLHKASKTKFTG-----YAGAV-----YDLNNDNSLYL-----SLSOLYTPQT 518
 Db 595 IRYDYRSTHSEDKSVSTGTHRNLSWNAQVVLKPFETWMDLTYRASTGPRLPSPFAEMYGWRA 654
 QY 519 NLDADGKLLKPRQNGQEVG--YKG-----SYMDDLNAARVSP-YRMKDKN-----AA 563
 Db 655 GESLTKLDLKPESFNREAGIVFKGDFGNLEASVFNNAAYRDLIAFGYETRTQNGOTSASG 714
 QY 564 APLNPNKKTRYAA---LGRVMEGVETEISGAVTPKW-----QIHAGYSYLHSGIKTAS 615
 Db 715 DPGYRNAQNARIAGINILGKIDWHGV-----WGLPDLGLYSTLAYNRKVKDA- 762
 QY 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 Db 763 DIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWINTMETYSKAKSVDELGSQLLN 822
 QY 647 GGVNAMSGITSSAGMHAGGYATFDMAAAYRFTPKLKLQINADNIFNRHY----- 695
 Db 823 GNANA-----KKAASRRTRPWYTVDSVGYNYIKKHLTLRAGVYNLLNRYVTWENVROTAG 878
 QY 696 -----YARVGGANTFNIPGSERTWTANLRY 720
 Db 879 GAVNQHKNVGVYNYAAPGRNYTFSLEMKF 908
 RESULT 42
 US-08-337-483-94
 ; Sequence 94, Application US/08337483
 ; Patent No. 5923562
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena
 ; APPLICANT: Harkness, Robin
 ; APPLICANT: Schryvers, Anthony
 ; APPLICANT: Chong, Pele
 ; APPLICANT: Gray-Owen, Scott
 ; APPLICANT: Yang, Yan-Ping
 ; APPLICANT: Murdin, Andrew
 ; APPLICANT: Klein, Michel
 ; TITLE OF INVENTION: Transferrin Receptor Genes
 ; NUMBER OF SEQUENCES: 147
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney

STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-410 MIS:jfb
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-337-483-94

Query Match 4.4%; Score 167.5; DB 2; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QFMSVPRIMTAATVLAALSSVFA-----AQTADLETVHLKGORVYAVITEKXGDY 55
DB 2 QOQHLEFRLILICSLMTAL--PYAENVQAEQOEKQDLTIQYKAKKQKTRRNEVTG-- 57
QY 56 SFEAVTVGKIPASLREIPQSVSIITNOQYKRNVDFTDQARKTPGLRYLNDGRSSV 115
DB 58 -----LGLVKSDDLKKEQVLN-----IRDLTYDQGLAVVEQGRGASSG 98
QY 116 YA-RGEYSESYN--IDGLPAQWOS-----INGTLP-----NLPAFDR--VEVWRCP 156
DB 99 YSIRGMDKRVSLTVDGV--SQIQSYTAQALGRTTAGSGAINEIEYENVKAVIISKGS 157
QY 157 SGLFSSGEMGGLV-----171
DB 158 NSEYENGALAGSVAFOCTTAADIIGEGKQWGIQSKTAYSGKDHALTOSIALAGSGGAE 217
QY 172 ---LVKRPPTKAFQGHAAAGFTH-----KQYK-----AEADVSG 203
DB 218 ALLIYTKRGRREIHAHKDNGKGVQSFNRLVLDDBKKEGSGQYRFVIEECHNGYAAKKN 277
QY 204 SLNSDSGV---KRVMAQTVGASPR---PAE-----KNNRH-----233
DB 278 KLEDEASVDERKTVTODYTGSRNLANPLEYSGSGLFRPGMHLNRYHVAVLERTQ 337
QY 234 ETF-----YAADWDINP-----DTVIGAGY-----LY 256
DB 338 QTFDTDMTVPAVFTSB-DVPSGLKGLKGYSGDNKAERLFVQEGSGTLOGIGTGCVFY 396
QY 257 QORHLAPYNGLPADANNKLPJSLPOHVFVGADWNKFKKNSHDFADLKHVFGNGGQYGVGM 316
DB 397 DERHTKRGV-----EYVYHND-----KDTWAD-----YARLSYDRGI 432
QY 317 RYSDRDADSNYAAGSKLGMKTPAGRGPCNTADDKACAVGLGEIHKOKALAPASYSRPF 376
DB 433 DLNDRLOQTHCSHDGSKNCRPDGNKPYSFYKSDRMT-----YEESRNLFOAVFKAF 485
QY 377 RLGNLANEFYIGADYRFRSTNGGRTLLYAFGGLALNFRSIPQVLLINAR---KGV 432
DB 486 DTAKIRHNLSINIGYDRFKS--QLSHSDYLLQ-----NAVOAYDLITPKKPPPPNGS 535

QY 433 RGYSHVA-----TENIDFCITGKSTFHPADGLSLIGG-----RLGHYKIESGEG 479
DB 536 KDNPFKVSIGKTTVTNSPICRFGNNITYTDCPTPRNIGGNGYAAVDNVRLAGRW-ADVAG 594
QY 480 -----KTLHKASKTKPTG-----YAGAV-----YDLNDNNSLYL---SLSQLYTPQT 518
DB 595 IRYDRSTHSEDSKSVSTGHRNLSMNAGVLLKPFWMMDLTYASTGFRLLPSFAEMTGMFA 654
QY 519 NLDADGLKLPKQNGQFEVG--YKG-----SYMDDRINAEVSP-YRMKDKN-----AA 563
DB 655 GSKLTKDLKPKERSFRRREGIVFKDGFGLNLSYFNNAARDLJARGYERTONGQTSASG 714
QY 564 APLNPNKKTRVAA---LGRVMEGVETEISCAVTPKW-----QIHAGSYHSQIKTAS 615
DB 715 DGYRNAQNRJAGINILCKIDMHGV-----MGCLPDGAYSTLAINRIKVKDA- 762
QY 616 NRDDGIFL-----LMRK-----HSANLW-----TTYQ-----VTPELTIG 646
DB 763 DLRADRTFTSYLPDAVQSRVYLGIDYDPDGIWGINMTFTYSKAKSVDELIGSQALLN 822
QY 647 GGVNAMSGITSSAGMHAGVATFDMAAAYRFTPKLKLQINADNIENRHY-----695
DB 823 GNANA-----KKAASRRTRPWYVTVDSGYINIKKHLTLRAGVYNLNLNRYVTWENRQTAG 878
QY 696 ----YARVGNANTENIPGSERTWTANLRY 720
DB 879 GAVNQHKNVGVNRYAPAPGRNYTFSLMKF 908

RESULT 43
US-08-478-373-94
Sequence 94, Application US/08478373
Patent No. 5922841

GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-ping
APPLICANT: Mardin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,373
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973

REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-478-373-94

Query Match 4.4%; Score 167.5; DB 2; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QPMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQORSYNAIVTEKNGDY 55
DB 2 QOHLFRNLICLSLMTAL--PYAENVOAQEQKQDITQVAKKQKTRDRDNEVTG-- 57
QY 56 SSFAVTGTGKIPASLREIPQSVSIITNOQVKDRNVDTFDLARKTFGLRVLSNDDGRSSV 115
DB 58 -----LGLKSSDTSLSKEQVFN-----IRDLRYDPGIAVVEQGRGASSG 98
QY 116 YA-RGYEYSEYN--IDGLPAQWQS-----INGTLP-----NLPAFDR-----VEVMRGP 156
DB 99 YSIRGMDKNRVSLTVDGV--SQIQSVTAQAALGGTTRTAGSSGAINEIEYENKVAIVEISKGS 157
QY 157 SGLFDSSGSGMGIVN-----SGLFDSSGSGMGIVN-----SGLFDSSGSGMGIVN----- 171
DB 158 NSSEYVNGALAGSVAPOTTAADIIGEGKQWGIQSKTAYSGKDHATOSLALAGSSGAE 217
QY 172 ---LVKRPTKAFQGHAAAGFGTH-----KQYK-----AEDVSG 203
DB 218 ALLIYTKRGRETHAKDAGKQVSNRLVDBDKKGGSQRYFIVEECHNGYAAACKN 277
QY 204 SLNSGDSV---RGRVMAQTVGASPR---PAE-----KNNRH----- 233
DB 278 KLKEDASVKDERKTVSTQDYTGNSRLLANPLEYSGSWLFRPGWHLDRHRYVGAVLERTQ 337
QY 234 ETF-----YAADWDINP-----DTVLGAGY-----LY 256
DB 338 QTFDTRDVTVPAYFTSE--DYVPGSLKGLKYGSDNKAERLFVQGGSTLQIGYGTGVFY 396
QY 257 QORHLAPYNGLPADANNKLPSPQHFVGVGADWNKPKMNSHDFADLKHYFGNGGKGYGM 316
DB 397 DERHTKNRYGV-----EYVYHNAD-----KDTWAD-----YARLSYDRQGI 432
QY 317 RYSDRDADSNYAFAGSKLGMKTPAGRPGCNTADDKACAVGLGTEIKQKALAFDASYSRPF 376
DB 433 DLNRLQOQTHCHSDGSDKNCRDPGNKPYSFYKSDRMI-----YEESRNLPAVFKKAF 485
QY 377 RLGNATANEPVIGADYNRFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANAR-----KGV 432
DB 486 DPAKTRHNLINSLNGYDRFKS--QLGSHSDYLLQ-----NAVQAYDLITPKPPFPNGS 535
QY 433 RGYSHTVFA-----TENLDFEIGYKSTFFHPADGLSLTGGG-----RLGHYKIESGEG 479
DB 536 KNPYRVISGKTTVNTSPICREGNNTYTDTCTPRNIGNGGYAAVQDNVRLGRW-ADVAG 594
QY 480 -----KTLHKASKTKPTG-----YAGAV-----YDLNNDNSLYL-----SLSQLYTPTQ 518
DB 595 IRYDYRSTHSEKSVSTGTHRNLSNAGVVLKPFPTMTDLYTASTGFRPLPSFAEMYGWRA 654
QY 519 NLDADGKLLKPPQGNQFVG--YKG-----SYMDRLNARVSP-YRMKDKN-----AA 563
DB 655 GESLTKLDLKPESKFNREAGIVFKGDFGNLEASYNFNAYRDLIAFGYETRTQNGQTSAG 714
QY 564 AFLPNNNKTRVAA---LGRKRVMEGVETEISGAVTPKW-----QIHAGVSYLHSGIKTAS 615
DB 715 DPGYRNAQWARIAGINILGKIDWHGV-----WGGLPDGLYSTLAYNRKVKDA- 762
QY 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTTELIG 646

DB 763 DIRADRTFVTSYLFDAVQPSRYVLGLGDYHDPDGIWINTMFTYSKAKSVDELGSQALLN 822
QY 647 GGVNAMSGITSSAGMHAGGYATFDAMAAYRFTPKLKLQINADNIFNRHY-----695
DB 823 GNANA-----KKAASRRTRPWTVDVSGYINIKKHLTLRAGVYNLLNRYVVTWENVROTAG 878
QY 696 -----YARVGAGANTFNIPGSRERTWTANLRY 720
DB 879 GAVNQHKNVGVYNYRYAAGPRNYTFSELMKF 908

RESULT 44
US-08-474-671-94
Sequence 94, Application US/08474671
Patent No. 6008326
GENERAL INFORMATION:
APPLICANT: Loosmore, Sheena
APPLICANT: Harkness, Robin
APPLICANT: Schryvers, Anthony
APPLICANT: Chong, Pele
APPLICANT: Gray-Owen, Scott
APPLICANT: Yang, Yan-Ping
APPLICANT: Murdin, Andrew
APPLICANT: Klein, Michel
TITLE OF INVENTION: Transferrin Receptor Genes
NUMBER OF SEQUENCES: 147
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,671
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/337,483
FILING DATE: 08-NOV-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/175,116
FILING DATE: 29-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,968
FILING DATE: 08-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-474-671-94

Query Match 4.4%; Score 167.5; DB 3; Length 908;
Best Local Similarity 18.2%; Pred. No. 1.8e-06;
Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QPMSVFRINMTAATVLAALSSVFA-----AQTADLETVHIKQORSYNAIVTEKNGDY 55

Dh	2	QOQH.FRNLN:CLSLMPLT--PVAAENVAQBQAEOKOLDITQVAKAKKOKRRRDNELVTG--	57
Qy	56	SSFAVTVGTAKLPASIRELPDSVSIITNQOYKDRNVDTFDOLAKTRPGLRVLSNDDGSSV	115
Dh	58	-----LGKLVKSSDLSLKEOVLN-----IRDLTRYDPGAIAVEOGRGSSG	98
Qy	116	YA-RGEYESEVN--IDGLPAMOS-----INGTLP-----NLFAFDR---VEVWRGP	156
Dh	99	YSIKGMDKNRSLVYDVG-SQIQSYTAQALMGIRTAGSSGAINETIYENVAKEISKS	157
Qy	157	SGLEPDSGSEMGIYN-----	171
Dh	158	NSSEYGNALGSVAFOPTKADIIEGQKWQIOKTAYSCKDHALTQSLALAGSSGAE	217
Qy	172	---LVKKEPTAFQGHAAAGFCH-----KOYK-----AENDVSG	203
Dh	218	ALLIYTKRRGRBIHAHKDAGGVOSFNLVLVDEDKKEGSGOYRPIVEBCHNGYAAKKN	277
Qy	204	SLSNGSV--RGHVMQTVGASPR---PAE-----KNNRH-----	233
Dh	278	KLKEDASVKDERKTVSTQDITYGSRLLANLPBYSQSQWLPFRPGHLDNRHYVGAVLERTQ	337
Qy	234	ETF-----YAADMDINP-----DTVLGAGY---LY	256
Dh	338	QTFEDRDMTVAYEAFYTS--DYVPGSLKGLGKYSGDNKAERLFVQEGSGSTLQIGIYGTEGVFY	396
Qy	257	QOORHLAPYNGLPADANNKLPBLPOHVFVGADPMNFKKONSHDVPFADLKHVYGGNGYGVGM	316
Dh	397	DERHKRKYV-----EYVHNAD-----KDTMAD---YARLSYDRQGI	432
Qy	317	RYSDBRADSNYAFAGSKLGMKTPAGRPQCNTADDKACAVGLGTEIKOKALAFDAYSRRP	376
Dh	433	DLDNRLOOHTCSHSGDKNCRPDGNKPYFSFKSDRMI-----YEESRMLFOGVFKCAF	485
Qy	377	RLGNANFVGLGADNVRSRSTNEGORTLYVARGGLALNEPSTIPOVULINAR---KGV	432
Dh	486	DTAKRNLINSLNGYDRKS--QUSHSYLYQ-----NAVOAYLITPPKPPFPNGS	535
Qy	433	RGYSHTVA-----TENLDEFGYGSTFHPDAGLSLIGGG-----RLGHYKIESGEG	479
Dh	536	KDNFYRVSIGKTVVTSPICFGNNTYIDTCPRNIGGNGYAAVQDNVRLGRW-ADVGAG	594
Qy	480	-----KTLHAKSKTYFGT-----YAGAV---YDLDNNSLYL---SLSQLYTPQT	518
Dh	595	IRDYRSTHSBDSKYSTETHRNLSMNAAGVVLKPTWMDLTLYRSTGFRLLPSFAEMYGMA	654
Qy	519	NULDADGKLLKPEROGQOFVWG--YKG-----SYMDDRILNARVSF-YRMDKN-----AA	563
Dh	655	GESLKTLDLKEKESNNRAGIVFPGDPEANLEASFPNNAYRDLIAFGYETRPONGQTSASG	714
Qy	564	APLNNNNKRTYAA--LGKMEGVETEISGAVTPKM-----QIHAGSYLSHQIKTAS	615
Dh	715	DGYNAQNAIRIAGINILGKIDMHGV-----WGLGPDGLYSTLAINRIKVKDA-	762
Qy	616	NSRDDGIGL-----LMPK-----HSANLM-----TTYQ-----VTPELTIG	646
Dh	763	DIRADRTVTVTLYEPDAVOPSRYYVLGLGYDHPDGIWGINMTFTYSKASVDELLIGSQALLN	822
Qy	647	GGVNAMSGITSSAGMHAAGYATFDPMAAYRTPPLKLIQINADNIENRHY-----	695
Dh	823	GNANA-----KKAASRRTPEWTVYDVGSYNKKGLTLRAGVYLLNLNRYTWENVRTAG	878
Qy	696	-----YAVGGANTPNITGSERTANLRY	720
Dh	879	GAVNQHKNVGVYNNRYAAGRYTTSLEMKF	908

RESULT 45
US-08-483-577A-94
; Sequence 94, Application US/08483577A
; Patent No. 6015688
; GENERAL INFORMATION:

```

1 APPLICANT: Loosmore, Sheena
2 APPLICANT: Harkness, Robin
3 APPLICANT: Schryvers, Anthony
4 APPLICANT: Chong, Pele
5 APPLICANT: Gray-Owen, Scott
6 APPLICANT: Yang, Yan-ping
7 APPLICANT: Murdin, Andrew
8 APPLICANT: Klein, Michel
9 TITLE OF INVENTION: Transferrin Receptor Genes
10 NUMBER OF SEQUENCES: 160
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Sim & McBurney
13 STREET: Suite 701, 330 University Avenue
14 CITY: Toronto
15 STATE: Ontario
16 COUNTRY: Canada
17 ZIP: M5G 1R7
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patentin Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/483,577A
25 FILING DATE: 07-JUN-1995
26 CLASSIFICATION: 435
27 PRIOR APPLICATION DATA:
28 APPLICATION NUMBER: US 08/337,483
29 FILING DATE: 08-NOV-1994
30 PRIOR APPLICATION DATA:
31 APPLICATION NUMBER: US 08/175,116
32 FILING DATE: 29-DEC-1993
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US 08/148,968
35 FILING DATE: 08-NOV-1993
36 ATTORNEY/AGENT INFORMATION:
37 NAME: Stewart, Michael I
38 REGISTRATION NUMBER: 24,973
39 REFERENCE/DOCKET NUMBER: 1038-511
40 TELECOMMUNICATION INFORMATION:
41 TELEPHONE: (416) 595-1155
42 TELEFAX: (416) 595-1163
43 INFORMATION FOR SEQ ID NO: 94:
44 SEQUENCE CHARACTERISTICS:
45 LENGTH: 908 amino acids
46 TYPE: amino acid
47 STRANDEDNESS: single
48 TOPOLOGY: linear
49 US-08-483-577A-94

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Inc.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:40:34 ; Search time 27.0814 Seconds
(without alignments)
2562.980 Million cell updates/sec

Title: US-09-889-267-2

Perfect score: 3776

Sequence: 1 MGQFMSVFRINMTATVLA.....NTFNIQSGERTWTNLRYSF 722

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3244	85.9	708	2 B81038	TonB-dependent rec
2	832	22.0	802	2 A83125	probable TonB-depe
3	828.5	21.9	826	2 S41569	ferric pseudobacti
4	822.5	21.8	815	2 A40601	ferripyoverdine re
5	821.5	21.8	815	2 H83345	ferripyoverdine re
6	788	20.9	725	2 D81976	probable ferric si
7	784	20.8	725	2 H81030	TonB-dependent rec
8	767.5	20.3	819	2 S15169	ferric-pseudobacti
9	755	20.0	753	2 AE0187	probable iron-side
10	737.5	19.5	729	2 C64854	ferric-coprogen re
11	736.5	19.5	729	2 H90813	outer membrane rec
12	734.5	19.5	729	2 D85673	ferric-pseudobacti
13	715.5	18.9	720	2 A36942	Fe(III)-pyochelin
14	708.5	18.8	809	2 S13899	ferric-pseudobacti
15	512	13.6	863	2 AF2074	ferrichrome-iron r
16	486.5	12.9	820	2 AE2130	ferrichrome-iron r
17	486.5	12.9	885	2 AB1944	ferrichrome-iron r
18	483.5	12.8	854	2 AE2082	ferrichrome-iron r
19	481	12.7	858	2 AD2421	ferrichrome-iron r
20	479	12.7	851	2 D87252	TonB-dependent rec
21	475	12.6	863	2 A12077	ferrichrome-iron r
22	470	12.4	760	2 E64817	probable membrane
23	467.5	12.4	867	2 AC2140	ferrichrome-iron r
24	465	12.3	828	2 S74450	ferrichrome-iron r
25	460.5	12.2	851	2 AE2129	ferrichrome-iron r
26	460	12.2	760	2 C90739	hypothetical prote
27	460	12.2	760	2 E85589	hypothetical prote
28	459	12.2	708	2 A83043	probable outer mem
29	450.5	11.9	747	2 G97408	fegA protein U6140

30	450	11.9	819	2 A13197	TonB-dependent rec
31	449.5	11.9	872	2 AC2134	ferrichrome-iron r
32	441.5	11.7	708	2 AH2626	ferrichrome-iron r
33	441.5	11.7	753	2 D83081	probable outer mem
34	433.5	11.5	732	2 A83481	probable TonB-depe
35	424	11.2	858	2 AE2085	ferrichrome-iron r
36	421.5	11.2	696	2 AC0547	ferrichrome-iron r
37	420.5	11.1	853	2 AC2079	ferrichrome-iron r
38	416.5	11.0	853	2 S74457	ferrichrome-iron r
39	415.5	11.0	857	2 AC2132	ferrichrome-iron r
40	411	10.9	714	2 C95382	probable ferrichro
41	407.5	10.8	635	2 C81861	hypothetical prote
42	402	10.6	863	2 S74447	ferrichrome-iron r
43	396.5	10.5	710	2 S22673	ferrichrome-iron r
44	396.5	10.5	802	2 C83588	probable hydroxama
45	391	10.4	713	2 B91118	probable ferrichro

ALIGNMENTS

RESULT 1

B81038

TonB-dependent receptor NMB1829 [imported] - Neisseria meningitidis (strain MC58 serogro

C/Species: Neisseria meningitidis

C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C/Accession: B81038

R/Tetreltin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.

Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;

Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masiarman, V.; Piza, M.

Science 287, 1809-1815, 2000

A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve

A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A/Reference number: AB1000; MUID:20175755; PMID:10710307

A/Accession: B81038

A/Status: preliminary

A/Molecule type: DNA

A/Restrictions: 1-708 <TET>

A/Cross-references: GB:AE002532; GB:AE002098; NID:G7227078; PIDN:AAF42164.1; PID:G722708

A/Experimental source: serogroup B, strain MC58

C/Genetics:

A/Genes: NMB1829

Query Match	85.9%	Score 3244	DB 2	Length 708
Best Local Similarity	86.8%	Pred. No. 6.1e-219		
Matches 627	Conservative 33	Mismatches 48	Indels 14	Gaps 3
QY	1	MGQFMSVFRINMTATVLAALSSVFAAQTADLETVHIGORSYNAIVTEKNGDYSSFAV	60	
DB	1	MGQFMSVFRINMTATVLAALSSVFAAQTADLETVHIGORSYNAIVTEKNGDYSSFAA	60	
QY	61	TGCTKTPASLRIPQSVSIITNOQVDRVDFPDQARTPGLRVLSNDGSSVARYQ	120	
DB	61	TGCTKTPASLRIPQSVSIITNOQVDRVDFPDQARTPGLRVLSNDGSSVARYQ	120	
QY	121	EYSEYVIDLPAWOSINGTLPNLFAFDREVMRGPSGLFDSGSEMGCIIVNVRKPTA	180	
DB	121	EYSEYVIDLPAWOSINGTLPNLFAFDREVMRGPSGLFDSGSEMGCIIVNVRKPTA	180	
QY	181	FOGHAAGFGTHKQYKAEADVSGSLNSDGSVGRVAAQTVGASPRPAEKNNRRETFYAA	240	
DB	181	FOGHAAGFGTHKQYKAEADVSGSLNSDGSVGRVAAQTVGASPRPAEKNNRRETFYAA	240	
QY	241	DMDINDPTVLGAGYLYQOCHLAPYNGLPADANNKLPSPLOHVFVGADWKKFKFNSHDVA	300	
DB	241	DMDINDPTVLGAGYLYQOCHLAPYNGLPADANNKLPSPLOHVFVGADWKKFKFNSHDVA	300	
QY	301	DLKHVYNGGNGYGVGRVYDRDADSNVAFAGSKLGWKTPEGPCNTADDKACAVGLGTE	360	
DB	301	DLKHVYNGGNGYGVGRVYDRDADSNVAFAGSKLGWKTPEGPCNTADDKACAVGLGTE	360	
QY	361	ITOKALAFDASYSRPFRLCNTANEFYTGADYVFRSTNEQSTTLVYARGGLALNERSLP	420	
DB	361	ITOKALAFDASYSRPFRLCNTANEFYTGADYVFRSTNEQSTTLVYARGGLALNERSLP	420	

Db 349 IKQKAFVADASYRPPALNTANEFVIGADYNLRSTNEQGRSTL--SKSVALDGFRLP 406
Qy 421 QVDLIANARKVGRVGSHTVATENLDFGIIYKSTFHPADGLSLGGRLGHYKIESGEGK 480
Db 407 YNGILQARAGNAGNFHNSVTEENLDETGLYAKTVFRLPGLSLIAGGRVGHKIESGDK 466
Qy 481 TLHKASKTKFTGYAGAVYDLNNSLYLSLSOLYTPQTNLDADGKLLKPRQNGQEVGYK 540
Db 467 TLHKASKTKFTGYAGAVYDIDGNSLYASASQLYTPQTSIGTDGKLLKPRQNGQEVGYK 526
Qy 541 GSYMDRLNARYSFYRMKDKNAAALPNPNKKTRVAALGKRVMEGVETEISGAVTPKQI 600
Db 527 GSYMDRLNTRVSYFYMKDKNAAALPDSNNKTRVAALGKRVMEGVETEISGAVTPKQI 586
Qy 601 HAGSYLSHISOIKTAGNSRDRDGLFLMPKHSANLWTTQVTPELTIGGGVNAWSGITSSAG 660
Db 587 HAGSYLSHISOIKTAGNSRDEGIFLLMPKHSANLWTTQVTPSLTIGGGVNAWSGITSSAG 646
Qy 661 MHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGGANTNIPGSERTWTANLRY 720
Db 647 IHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGSSESTNIPGSERTWTANLRY 706
Qy 721 SF 722
Db 707 SF 708
RESULT 2
A:Title: TonB-dependent receptor PA4168 [imported] - Pseudomonas aeruginosa (strain PA01
A:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83125
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; UID:20437337; PMID:10984043
A:Accession: A83125
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-802 <STO>
A:Cross-references: GB:AE004833; GB:AE004091; NID:g9950370; PIDN:AAG07555.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
C:Gene: PA4168
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homc
Query Match 22.0%; Score 832; DB 2; Length 802;
Best Local Similarity 30.0%; Pred. No. 4.5e-50;
Matches 222; Conservative 137; Mismatches 309; Indels 72; Gaps 21;
Qy 19 AALSSSVFAAQTADLETVHIKG-QRSYNAIVTEKNGDYSSFAVTVGTIKIPASLRIPQSV 77
Db 97 AATGAGIAPGSDVLDGALSVATQDPATAITFEDSGSYTTRAMRTSTKLAMSIRETPQSV 156
Qy 78 SIITNQVKDRNVDFDQLARTPGRLVLSNDGGRSSVYARGYEYSEYNIDGLPAQMOSI 137
Db 157 SVVTRQRMDDQGRDLNDVAKGVGTGLTVQOYGPARKVYAARGFDVDNIMYDGLPTSISTY 216
Qy 138 NG---TLNPLFAFDRVEMVRGSGFLFDSGSGGGIYNLVRKRPYKAFQGHAAAGGTHKQ 194
Db 217 TQDVIISAADLAMDFRVVEVVRGATGLMQAGNPAALINMVRKRPYKAFQGHAAAGGTHKQ 276
Qy 195 YKAEADVSGSLNSDCSVGR-VMAQTGASPRPAEKNRHETFYAAADWDINPDITVLGAG 253
Db 277 YKAEADVSGSLNSDCSVGR-VMAQTGASPRPAEKNRHETFYAAADWDINPDITVLGAG 253
Qy 254 YLYQQRHLAP-YNGLPADANNKLPSPQHFVFGADWNKFKMNSHDVADFADLKHFGNGGYYG 312
Db 336 ASNQNDNRNDNWVGLPSPGGRHLDLKRSSYYGADSWYDWDITTHLFDLTHRFANGWQM 395

Qy 313 KVMRYSDRDADSNYAFAG-----SKLGMKTPAGRPG-CNTADDKACAVGLGTEI 361
Db 396 KLA-----ADKLWAINMLGLYNDYCYSTTGCSMTQNPQDYSYTDHHD----- 439
Qy 362 KQKALAFDASYSRFRICNTANEFVIGADYNLRSTNEQGRSTLTYARGGL---ALNEFRS 418
Db 440 -----SYDAYANGFPQLLGRHEHLVVGASVQERFDGHHGWSLFPKDGTPGMDPTQM 494
Qy 419 IPOVDLIANARKVGRVGSHTVATENLDFGIIYKSTFHPADGLSLGGRLGHYKIESGE 478
Db 495 DPSSTLKPRLNTSLWGMK-----LDQEQKGYALTTRNLADPLKVLGRLDWMYKADAT 549
Qy 479 KTLHKASKTKFTGYAGAVYDLNNSLYLSLSOLYTPQTNLDADGKLLKPRQNGQEVGY 538
Db 550 DS--YKVTN-NTVTRVAGVIYDLNQTYSVYATDIFKPSQNFAGGGLDITKQNYEIG 606
Qy 539 YKGYMDRLNARYSFYRMKDKNAA-----APLNPNKKTRYA--ALGKRVMEGVETE 587
Db 607 LKGFHFGALNSQIALFOIDENRATEDVGGPSPCFPSPT---SRYCSRASGKVRSGQVD 663
Qy 588 TEISGAVTPKQIYHAGSYLSHISOIKTAGNSRDRDGLFLMPKHSANLWTTQVTPEL-- 643
Db 664 LELSGALSDDWQMWAGTYVDKVKHDSNKANECKPFDAAKPRHLFKLATSYTLPGLHLK 723
Qy 644 -TIGGGVNAWSGI-TSSAGHHAGGYATFDMAAAYRFTPKLQINADNIFNRHHYARVGG 701
Db 724 WRVGGDLATOSKTESDSTGFGQGGTYVVMALGYKVNERIDTRLNPNFLDKKYYSIDF 783
Qy 702 ANTENIPGSERTWTANLRY 721
Db 784 GN-LNY-GEPRNLMTVKYS 801
RESULT 3
A:Title: TonB-dependent receptor protein - Pseudomonas sp.
A:Species: Pseudomonas sp.
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 08-Oct-1999
C:Accession: S41569; S34357
R:Morris, J.; Donnelly, D.F.; O'Neill, E.; McConnell, F.; O'Gara, F.
Mol. Gen. Genet. 242, 9-16, 1994
A:Title: Nucleotide sequence analysis and potential environmental distribution of a ferr
A:Reference number: S41569; UID:94104606; PMID:8277948
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-826 <MOR>
A:Cross-references: EMBL:X73412; NID:g313725; PIDN:CAA51812.1; PID:g313726
C:Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homc
F:183-317/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:540-826/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
Query Match 21.9%; Score 828.5; DB 2; Length 826;
Best Local Similarity 29.9%; Pred. No. 8.4e-50;
Matches 227; Conservative 129; Mismatches 303; Indels 101; Gaps 26;
Qy 16 TVLAALSSSVFAAQTADLETVHIKGQRSYNAIVTEKNGDYSSFAVTVGTIKIPASLRIPQ 75
Db 115 TVLGPATGSAM-----ELAPTNVNASRL--GATTEGNSYTTGGVTIGKV-HSLKETPQ 166
Qy 76 SVSIITNQVKDRNVDFDQLARTPGRLVLSNDGGRSSVYARGYEYS-BYNIDGLP--- 131
Db 167 SVTVNTRKMLDDQNLNLTIEQVMEKTPGITTVDSPMGKGYFSRGRFMSGGQYQYDGVPLDI 226
Qy 132 ----AQMOSINGTLNPLFAFDRVEMVRGSGFLFDSGSGGGIYNLVRKRPYKAFQGHAA 187
Db 227 GSSVYQADSFNS---DMAIYDRVEVLRAAGMMKAGGTAGGVNFRKRGQDTAHTQLSL 283
Qy 188 GFGTHKQYKAEADVSGSLNSDCSVGR--VMAQT-----VGASPRPAEKNRHETFYAA 239
Db 284 SAGTMDNRYGQVDTGGFLNSDGTIRGAVTEQTRQYFYDVG-----SRKQDIYYGA 335

Query Match 21.8%; Score 822.5; DB 2; Length 813;
Best Local Similarity 28.9%; Pred. No. 2,1e-49;
Matches 215; Conservative 145; Mismatches 317; Indels 67; Gaps 23;

```

OY      8 FRIMNTATVLAALSSSFAAQTADLETVHIKGRSRNAIYTEKNGDVSFAVTGKIP 67
Db      108 PQGNATITTSVEADSSV-----DLGATMTTSNQL--GITTEDSGYTPECTIATRLV 159

OY      68 ASLNEIPQSVAILITNOQYKDRNVDTPDLARKTPGLRYLSNDDRSSVYAGYESEYNI 127
Db      160 LTPRETPQSIIVTVRRQNMDPFGANNIDVMEHTFGITVSAVDTPBRNNVYAGFSINNPOY 219

OY      128 DGLPRAQOSI---NGTLPNLPAFDREYVMRGPGLFSDSGEMGIVNLVKRPPTKAFQG 183
Db      220 DGIPTANVVSAGANTLSDMAIYDRVEVLKGAITGLLTGASLTATTIILLIKKPTHEKRG 279

OY      184 HAAGFGTHKQYKAEADVSSGLNDDGSVRGVMQYTGASPRAPKNNRHETFFAAADD 243
Db      280 HVELGAGSWNDYRSELDVSGPLTESGNVRGAVAAYODKHSFMDHYERTSYVYGILEFD 339

OY      244 INPDTVLGAGLYLQOORHL--APYNG-LPA--DANKKLPSLPQHEVFGADMNFKKXNSHDVF 299
Db      340 LNPDTMLTVGADYQDNDPKSGSGSFPFLPDSQGNRNDVSSFPNNGAKMSSWEDQYTRVF 399

OY      300 ADLKHVFGNGGYKGVKMKRYSDRDADSNVAFAGSKLGMKTPAGRPGCATDADKACAVGLGT 359
Db      400 ANLEHNFANVGWGVKQQL---DHKINGYHAPLGAIMG--DWPA-----PDNSAKIVAQKY 448

OY      360 EIKQKALAFDASYSRPPFLGNTANEFYIG--ADNRPFRSTNGEGRITLYIYAG--GLALNEF 416
Db      449 TGETKSNLIDLYLTGPPQFLGREHELVGTSSAFSHW----EGKSYWNLNLYDNTTDDF 503

OY      417 RSIPOVDIAARAKGVRSYHTVATENLDEFGIYKSTFPHADGLSLGGRLAGHYKIES 476
Db      504 -----IMNDGIGIKRPDWCTPSQYIDDKTRQJGSIYMTARFNTDNLNLFGRVVDYR-T 557

OY      477 GEGKTLHAKSKTKFTGVAAGAYVDLNDNNSLYLSLSQLYTPQTN--LDADGKLLPRQGNQ 534
Db      558 GLNPTRESG--RFIPYGAAYDLNDTFYSVAYATDIFMPQDSWYRDSNKLLEPDEQGN 615

OY      535 FEVGYKSGYMDRLNARSRFPMKDKNAAP----LNPNNKTRRYALG--KRWMEGYET 588
Db      616 YEIGIKGKYLDGRNNTSLAYEPIHEENRAEDALYNSKPTPAITVYAKGIKAKTKGYEA 675

OY      589 EISGAVTPEKMOIHAGYSYLSHQIKTASNRDDGIFLMPKKSANIMLTYYQ---VTPELTI 645
Db      676 EISGELAPGWQVQAGYT--HKIIRDSGKK--VSTWEPQQLSLFISYKRGKALDKLITV 730

OY      646 GGGVANAAGSITSSAG-----MHAGYATPDANAAYRFPKTLKLOINADNIENRHYAR 698
Db      721 GGGAARWQKSGNQMNPRRMWEKFSQEDYWLVDLMARVQITDKLSASVYVNVNVPKTYTIN 790

OY      699 VGGANTFNIPGSERTWTANLRYSF 722
Db      791 IGFYTSASY-GDPNLMFSTRWDF 813

RESULT 5
H83345
Feature: overdrive receptor PA3398 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H83345
R:Stover, C.K.; Yam, X.Q.; Errin, A.L.; Micooguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br-
adman, S.; Yuan, Y.; Brody, L.U.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Dim-
.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: H83345
.;Status: Preliminary
.;Molecule type: DNA
```


Query Match Similarity 20.8%; Score 784; DB 2; Length 725;
Beet Local Similarity 29.4%; Pred. No. 8,9e-47;
Matches 228; Conservative 118; Mismatches 320; Indels 110; Gaps 21;

RESULT 8

S15169
ferric-pseudobactin receptor precursor - Pseudomonas putida
C|Species: Pseudomonas putida
C|Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 29-Sep-1999
C|Accession: S15169
R|Biller, W.; Marugg, J.D.; de Weger, L.A.; Tommassen, J.; Weisbeek, P.J.
Mol. Microbiol. 5, 647-655, 1991
A|Title: The ferric-pseudobactin receptor PupA of Pseudomonas putida WCS358: homology to
A|Reference number: S15169; MUID:91260449; PMID:1646376
A|Accession: S15169
A|Molecule type: DNA
A|Residues: 1-819 <Bit>
A|Cross-references: EMBL:X56605; NID:g45722; PIDN:CMA9942.1; PID:g45723
C|Genetic8:
A|Gene: pupA
C|Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homo
C|Keywords: membrane protein
F.1-47/Domain: signal sequence #stratus predicted <SIG>
F.48-819/Product: ferric-pseudobactin receptor #stratus predicted <MAT>
F.189-322/Domain: tonB-dependent receptor amino-terminal homology <TN>
F.542-819/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 20.3%; Score 767.5; DB 2; Length 819;
Best Local Similarity 27.8%; Pred. No. 1.5e-45;
Matches 214; Conservative 134; Mismatches 306; Indels 117; Gaps 26;

OY	8	FRIINATVLAALSSVEAAQTADLETTHIKGRSYNAVTEKNGDYSSFAVTGKIP	67
Db	110	YQIDNTVTVASA-----AAKDQIELSATNVNSAGLEETGETSGATTFTSTATKM	164
OY	68	ASLRIPDSVGLITNQVKDKNVDTPQLAKRTGLRLSNDG-RSSVARGVEYEYN	126
Db	165	LISIRTPGITVVTRQRDDHDLGSMEVLTYTGFI-TMSODGEREFTYSRGSAINIYO	223
OY	127	IDGLPA----OMOSINGTLPMLFADREVNVRGSGFLFDSGEMGAIYVALRKSPPTAFQ	182
Db	224	FDGVTTYDNDOTRRNPSTILMDVGLYDRILEIVRGATGLMTGACDSAYVNVTRKKPTEFK	283
OY	183	GHAAGFGTHKOYKAADVSGSLNSDSGSRVNAQTVGASPRPAEKNNR-----H	233
Db	284	SHIQGVGSWPDYYAEADVSGPLTDDGRVGRFPA-----AKDNHTFMWDYTQDR	334
OY	224	EFTVAADMVINPDTVLGAGLYVQQRHLPYNGIPA---DANKKLPSLPQHVVGDWN	289
Db	335	DVLGVGVADVTDITVARFGI--DRQTYKVNGAPGPRIITYNQOPTFSRSTSSDARW-	390
OY	290	KPKMNSHVVPADLKHYFP-----NGGYGKVGMRSDDSDANVAFAGSKLGMTKPGRP	343
Db	391	----GYDDYTTNYTFLEEQALHDWOFKLAAYIMDVDRSESSY-----YSTTNR	439
OY	344	GCNTADKACAVGLGTBIKOKALAFDASYSRPFRLGNATANFVIADYNNERSTNEOGR	403
Db	440	YLELDGSTIEISAGIYT-AKHQHQGVDAITLQGFOLLGQTHELIYGVNVLEYENKH----	493
OY	404	TLYVAGGLALNEFRSIPOVDL-----IANARKG---VRGYSHTVATENIDEFGIYGKS	453
Db	494	----RGD-----SGRPVINIFYDMWDNQTKPPGDEIIPIGIYNISR--QGSGYFAS	539
OY	454	TFHPADGLSLIGGGRLGHYKIE-----SGEGKTLLHKASKTKFPGYGAAYVDLNDNNSLY	507
Db	540	RNLTDIDLHLTGABASVRFDPYALMRI GNBPARYKMYERGVVTPYAGIYIDLTNBSOV	599
OY	508	LSLSQLYTPQTNLADGKLLKPRQNOFEVYKGSYMMDDRANAVSFYRMKDKXAAA PLN	567
Db	600	ASYTDIFPKQNNDITGKPLDPVGVKNVELGMKEPLEGRLANNMIALYMVRKDLASTN	659
OY	568	---PNKKTRYAALGKRVMEGERT---EISGAVTPKMQIHAGSYLHSOIKTJANSRD	619
Db	660	EWVPSG---GLASRAVDAEFKGVDVLELGEVLPEMNVFCTYS--HTTEDADGKR-	712

Best Local Similarity 27.3%; Pred. No. 1.6e-43;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;

```

0Y 3 QEMSAFRINTMAATV-LBALSSVFAAQTADLETHIKGGSYNALIVEXKGDYSSFAVT 61
Db 11 QYALTKSSLLAGCIALALLPSPAFAAPATE-ETIVIVGSAV---APDGDENDSVSTIS 66
0Y 62 VQTKIPASLREIPQSVSIIITNOQVDRNVDPFDQJARKTPGLRVLSNDGSSVARSYE 121
Db 67 AGTKMQMTQRDPQSVTVIVSQGRMEDQOLQTLIGRMENTLIGSKQADSDRALYYSKRFQ 126
0Y 122 YSEYNIDGLPAQMOS---INGTLFNLFAFDRYEVNRGSPGLFDSSGEMGIVNLVKKRPT 178
Db 127 IDNYVWDGIPYTFEESRMNLGDALSDMALFEYREVVRGATGMLTGTGNPSAAINMVRKAT 186
0Y 179 -KAPQGHAAAGGTHKQYKAEADVSGSLNSDGSVARGMAQTGVA---SPRAEKNNEHE 234
Db 187 SREFPGDVSAEAGSNKKRYADLOSPLTEBCKIRARL---VGGYQNDSDMLDRYNEK 242
0Y 235 TFYAA-ADMDINPDPVLVAGLYYQOORHL--APYNGLP-A-DANNKLPJLPOHVFVADANXK 290
Db 243 TTFPGSVIADLDGLDTLLBSAGYEQKIDVNSPTWGLPFMRNTDGSNSIDRARSTAPMAY 302
0Y 291 FRKNSHDVPADLKHFGNGGYGKVMRYSDBDADSNT---AFAGSKJGMKTP----- 339
Db 303 NDKEINKFYMTLKQGFADTWQATLNAHSEVEFDSKMMYVAYVNAKADGMLVGPYSNGP 362
0Y 340 ----AGRGCTNADKACAVGLGTE-----IKQALALFADSYSPFR--LGNTNFR-- 385
Db 363 GPDVYVGGTGMSGKKRVADLDLFADGSYELFGROHNLFFGSGKONNRFFSSWANIIPD 422
0Y 386 VIGADYNFRSTNEQGTTLVARGGLALNEFRSIPQVDLIANARKGVRGSHVTATEULD 445
Db 423 EIGSPYN-FNG-----NFPQTD-----WSQSLAQDDTT 450
0Y 446 EF-GIYGKSTFHPADGLSLIGGRLGHYKISGEGKTLHKASKTYFTYGAAYVDLNDNN 504
Db 451 HMKSLYAATRVTLADPLHLILGARAYTNMRVDT---LTYSMENKHTTYPAGLVDFIDNMW 506
0Y 505 SLVLSLSQLYPQTNLDMDGKLLKPRQGNQPEVGYKGSYMDRLNARSFPRMCDKNA- 563
Db 507 STVASYTSIFQONDRDSSGKYLAPITGANNYELGKSDMMNSRLTTLTAIFRIEDQNVAQ 566
0Y 564 ---APLNPNKKTRPAALAGKRVMEGVETIEISGAVPQKQIHAGYS-YLHSQIKTNSGRD 619
Db 567 STGTFIPSGNSGNTAKAVDGYVSKVEFELNCAITDNNQJLTFGARIRYI-----AENDEG 620
0Y 620 DGIELMPKHSANLWTYQ--VTPELITGGGVNAMSGITSSA-----GHHAGYATFD 670
Db 621 NAVVNNLPRTTYKMFSTYRLPVMPELITVGGGVNQNRVYTDVTYVGTFRABQGSYALVD 680
0Y 671 AMAAARFPYKLTQJNADNIPFRHHYARVVGANTNIGSEERTYANLARYSF 722
Db 681 LFTROVTKNPSLQGNVNNLDPKTYDINVEGSIY---GTPNFSITGTYOF 729

```

C:Genetics:
A:Gene: EGS1480
C:Superfamily: ferric-pseudobactin receptor: tonB-dependent receptor amino-terminal hom

Query Match	Score	DB 2	Length
19.5%	736.5		729

```

0Y 3 QMSVSRIMNTATV-LAALSSSVFAAQADLETVHKKGRSYNATVTEKXGDYSFPATV 61
Db 11 QYQATKPKPBLLAGCICALLPASAFAAPATE-ETVIVEGSAT---APDGDENDSVTSTS 66
0Y 62 VGTKIPASLEIREIPOSVSIITNOOVKDRNVDPDQIAKTPCLRLVLSNDGSSVARGE 121
Db 67 AGTKMQMTQRDIPQSITVISOQRMEDQOLTLGEMENTLGISGSQADSDRALVYSRQFO 126
0Y 122 YSEYNIJLPAQOMOS--INGTLPLNLPAPDRVEVMRPSGLFDSSGEMGIVNLYRKRP 178
Db 127 IUNYWDGIPYTESHWNLDALSDMALFERVEYRGATGIMTGTGNPSAAINVYRKAT 186
0Y 179 -KAFQHAAAAGPETHQYKRAEADVSGSLSDSGSVRGVMAQTVCA---SPRPAEKNNHE 234
Db 187 SREFKDVAEAYGSMNKERYADLQSEPLTEBDGKIRARI---VGGYQNNDSWLDREYSEK 242
0Y 235 TFYAA-ADMDINDPYLAGYLYQOORL--APYNGLPA-DANNKLPSPLOHFVGDAMWK 290
Db 243 TFFPSGLVDADLDGLTLTSLAGYEYQRIIVNSPTWGLPRMNTDGSNSITDRARSTAPMAY 302
0Y 291 FKNNSHDFADLKHYPGNGGYGKVMRYSDRDADSNY---APAGSKLGKMT 339
Db 303 NDKIENKVFEMTLKQRPADWTQATLNTATHSEVEFDSKMMYVDAVYNNKADGMLVGYPYNYGP 362
0Y 340 ---AGRECNRAADPAACAVGLGTE-----IKXALAFDASYSRPRR--LGANTNEF-- 385
Db 363 GPDYVGGTWNSGSKRYVADALDLFADGSYELFGRQHNLMTFGSGYSKONNRYSSWANTIPD 422
0Y 386 VIGADVNRFRSTEGQRTLLYARGGLALNEFRSIPOVDLIANAERKVGYSHTVANTEND 445
Db 423 EIGSFYN-FNG-----NFPQPD-----WSPQSLAQDDTT 450
0Y 446 EF-GIYGKSTFHDADGLSLIGGRLGHYKLESSEGKTLHKASKTKFTGYAGAVYLDNN 504
Db 451 HMKSLYAAATRVTLADPLHLILGARVYTWBVDT---LTVSMERKHTTPYAGLVPDINDW 506
0Y 505 SLVLSLSQIYTPQTNLDADGKLLKPROGNOPEVGYKSSYMDRLNARVSFRMKCKMAA- 563
Db 507 STYASVTSIFOPONDRSDSGKYLAPITGNVNEYELGIKDDWMSRLTTTLTAIFRIEDDNVAQ 566
0Y 564 ---APLPNNKKRRYAALGRVMEGVEYETISGAVTFPMQIAGVS-YLHSQIKTASNSRD 619
Db 567 STGTLPDGSNGEYAYAAVDGTVSKGYEFELNGALTDMQOLTFGATRYI-----AEDNEG 620
0Y 620 DGI FILMPKHSANLMTTYQ--VTPELTI GGAVNAMSGITSSA-----GHHAGGYATFD 670
Db 621 NNVNPNLPRTYVMKFTSYRLPVMBELTVGGVVMQNNRYVDTVPYGFRAEQGSYALVD 680
0Y 671 AMAAIRFTFKLKQIYADNIFNRHYARVGCANTFNIPGSRRTTANLRYSF 722
Db 681 LFTRYOVNTKNSIQGNVNNLFDKTYDNTNVEGSIYV---GAPRNSITGTGYOF 729

```


Qy 651 AMSGIT---SSAGMHAGVATFDMAAAYRTPKLIKQINADNIENRHYARVGGANTENT 707
Db 646 AQGDYVDYRGVSMROGGYALVNMRLGYKIDEMHTAAVNVNMLFDRTYYQSLSPNPMNNR 705
Qy 708 PGSERTWTANLRYSF 722
Db 706 YGEPRSFVNSLRGAF 720

RESULT 14

S32899
ferric-pseudobactin receptor pubB precursor - Pseudomonas putida
C/Species: Pseudomonas putida
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #ext_change 29-Sep-1999
C/Accession: S32899
R/Koster, M.; van de Vossenberg, J.; Leong, J.; Weisbeek, P.J.
Mol. Microbiol. 8, 591-601, 1993
A/Title: Identification and characterization of the pubB gene encoding an inducible ferric-pseudobactin receptor pubB precursor
A/Reference number: S32899; MUID:93316856; PMID:8392140
A/Accession: S32899
A/Molecule type: DNA
A/Residues: 1-809 <KOS>
A/Cross-references: GB:X73598; GB:S63481; NID:9403024; PIDN:CAA51995.1; PID:9581467
C/Genetics:
A/Gene: pubB
A/Start codon: GNG
C/Superfamily: ferric-pseudobactin receptor; tonB-dependent receptor amino-terminal homology
C/Keyword: membrane protein
F/1-45/Domain: signal sequence #status predicted <SIG>
F/46-809/Product: ferric-pseudobactin receptor #status predicted <MAT>
F/189-319/Domain: tonB-dependent receptor amino-terminal homology <TNB>
F/543-809/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 18.8%; Score 708.5; DB 2; Length 809;
Best Local Similarity 25.7%; Pred. No. 2e-41;
Matches 197; Conservative 154; Mismatches 314; Indels 101; Gaps 23;

Qy 8 FRINMTAAVLAAL-----SSSVFAAQT-----DLETVHIKQGRVNAIYTEKNGD 54
Db 94 FDDIDGAIILLAGTGLASGANAISYLSQASGALFELASVSISSGAP--GSTTEGTGL 151
Qy 55 YSFAVTVGFKIPASLREIPQSVSIITNQVKNRVDTFDQLARKTGLRVLSNDG--R 112
Db 152 YTYSSSSSTRNLNTPRETPQSLTVMTRQRDLNLTALDLEATPGITVVRDGLSSES 211
Qy 113 SSYARAYESENIDLPAMQOSINGTLPNLPFDRVEVWREGSGLPDSSGEMGIVNL 172
Db 212 DSVMSRGFALQNYEVDPVPTSTR-LDNYVSQSMAMFDEIVEIRGATGLISGWNPSATINL 270
Qy 173 VRKRPRTAFPGHAAAGFTHKQYKAEADVSGSLNSDGSVGRVVAQTVGASPRPAEKNNR 232
Db 271 IRRKPTAEAOASITGEAGNMDRGTGFDVSGPLTETGTNRKRFVADYKTEKAWIDRYNQO 330
Qy 233 HETFYAADNDINDIVLAGAYLQQRHL-APY-NGLPAD-ANNKLPSLPQHFVGVADWN 289
Db 331 SQLMYGTTEBDLSDTLTVGFYSLRSDIDSLRSGLPTRSGTERNLKSLAAADWS 390
Qy 290 KFKKNSHD---VPADLKHYRNGGYGYKMYSPRDADSNAPFAGSKLGMKTPRAGPCGN 346
Db 391 ---YNDHEQTSFFTSISCOQNGMSGKIELTHANKFDELNFNA---MGEIAMPD----- 438
Qy 347 TADDKACAVGIGTEIKKALAF-----DASYSRPRLGNANFEVIGADVNRPRST 397
Db 439 -----GSGLSQLPVAFSGTSPRODNLILYATGPFSLPGEHEHLITMTLSQYREN 487
Qy 398 NEQRTTLVARGGLALNEFRSIPVD-----LIANARKGVRSYSHVATENLDERGIY 450
Db 488 TSPWGMGRDYAGSPAGAILNLFNMDGSAKAPAVESGKSI-----DDQYAAV 537
Qy 451 GKSTFHADGSLIGGRLGHYKTES-----GEGKTLHKAASKTKFTGYAGAYVDLNDNN 504
Db 538 LTRFSVYTDLSLILGRLINMKRDTSDRPYGEETEVRNRENGCVFLPYAGVGYDLDDTW 597

Qy 505 SLVLSLSQLYTPQTN--LDADKLLKPRQNGQEVGYKSIIMDDRLNARVSPYRMOKKA 562
Db 598 SLVASYTKLIFNPGCAVWTDENKPLDPMQGVGYELKIGTHLNGKLNSSIAVFKLQODNL 657
Qy 563 AAPLNNNKKTRYAALGXKVMGEVTEISGAVTPKKQIHAGVSYLHSQLKTSNSRDDGI 622
Db 658 ALWQHN---VYSAQODTTSKQIELELNGELBEGWQASAGYSY-----SVTTDADDQRI 708
Qy 623 FILMPKSNLMTTYOV---TPELTIGGVNAMSIGTSSAGMHA---GGYATFDMAAAYR 676
Db 709 NTMLPRANSFKTFTSYRLHGPDLKITIGGVNMQSKV--GADLHTFSGGSYAIVNLMARVD 766
Qy 677 FTPLKLIKQINADNIENRHYARVGGANTNIPGSERTWTANLRYSF 722
Db 767 INQHLASVNLNVFDPREYVSGG---LVGVGTPRNVMTSPFKYSF 809

RESULT 15

AF2074
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C/Species: Nostoc sp.
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #ext_change 30-Jun-2002
C/Accession: AF2074
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A/Reference number: AB1807; MUID:21595285; PMID:1175840
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-863 <KUR>
A/Cross-references: GB:BA000019; PIDN:BA073847.1; PID:917131239; GSPDB:GN00179
A/Experimental source: strain PCC 7120
C/Genetics:
A/Gene: all2148
C/Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homology

Query Match 13.6%; Score 512; DB 2; Length 863;
Best Local Similarity 25.0%; Pred. No. 1.3e-27;
Matches 182; Conservative 122; Mismatches 303; Indels 122; Gaps 28;

Qy 47 IYTEKNGDYSSFAVTVGFKIPASLREIPQSVSIITNQVKNRVDTFDQLARKTGLRVL 106
Db 204 VITGQONGVAVQDATTATKTIDPLRDIPOSIQVPRVLEDRVRSIAEAVTVSG--VV 261
Qy 107 SNDDGRSS---VYARGYE-----YSEYNIIDGLPAQMSINGTLPNLPFADRYEVMR 154
Db 262 DGADVNGSPAQDFLNGFEGGSGFRNGYRDVNSYGLTV-----ERVVLK 310
Qy 155 GPGGLPSSGEMGIYVNLAKRPTKAFQGHAAAGF--GTHKQYKAEADVSGSLNSDQSVR 212
Db 311 GPASVYLFQAVEPQGIINVTQKPLS--EPYQOLGFEGVGNARFQPSIDFGSLPADTTL 368
Qy 213 GRVWAQTVGASPRPAEKNNHETFYAAADWDI--NPDTVLAGAYLQQRHLAPYNGLPAD 270
Db 369 YRFNASTQSSDGDQDFVNTNLITIAPIAMKLDGRDRLITYEYINKKGFQDYTSLSD 428
Qy 271 ANNKLP-----SLPHVFGVADNNKFKMNSHDVADLKHYFGNGGYGKGVKMYSDR- DAD 324
Db 429 -NFTLPSPFYQAVYDNNAVVNTQKL-----GY-TLSHKFSMDNQIR 468
Qy 325 SNYAFAGSKLGMKTPRAGPCNTAADKACAVGIGTEIKKALAPDASYSR----- 374
Db 469 NNFSVVTSK-----NABEYTLATGVNQSLSQFQADREFTDQNNYFGQIDLL 515
Qy 375 -PPLIGTANAFVIGADYNNFRSTNEQRTTLVARGGLALNEFRSIPQVLIANARKGV 433
Db 516 GKREYVGSISHQIILGDFNNHIDT-----FANVQVRNVNLD-IRNPNINIP 561
Qy 434 GYSH---TVATENLDERGIYKSTFHPADGLSLIGGRLGHYKIESG--GKTLHKAASK 487

Db 562 SPDYGRSSSTERFOTYGIYLOQDTFLDNKLLIGRFDWISGENTDNVTGDTITQNPDS 621
Qy 498 TKFTGYAGAVYDLNNDNSLYLSLYTPQTNLDADGKLLKPRQGNQPEVGYKGSYMDDR 547
Db 622 SAFSPRIGLVYQPSKVSLSYTSYSFVPETGVNPDGEIFETRTQVEAGIKADFLSGR 681
Qy 548 LNARVSFYRMKDKNAAAPLNPNKTRVAAL-----GKRVMEGVETISGAVTPKQIH 601
Db 682 LSATLAAVQITKSNILTP-DPDPER---AALDYLIQVGEQRSGRIELDVAGBILPGWKAI 737
Qy 602 AYSYLHQSQIKTASNSRDDGIFLL-MPKHSANLWTTVQVTP-----LTIIGG---VNAMS 653
Db 738 ASYAYTNAEV-TEDNDIPVGNRLVSPKQASLWTTTFONSDKLGFLGFLVYVGTGRS 796
Qy 654 GITSAGMHAGYATFDMAAAYRFTPKLQINADNINRHYARVGGANTFNIGSERT 713
Db 797 G-DSANSFEIPDLRTDAIYR-RDGKAGINIRNLFDTDIRSDGRTFLRGAPFT 854
Qy 714 WTANLRYSF 722
Db 855 IIGSISWEP 863
RESULT 16
AE2130
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2130
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-820 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074295.1; PID:gl7131689; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2596
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.9%; Score 486.5; DB 2; Length 820;
Best Local Similarity 24.6%; Pred. No. 7.1e-26;
Matches 185; Conservative 112; Mismatches 353; Indels 103; Gaps 24;
Qy 15 AFVLAALSSSVFAAQ-----TADLETVHIKQRS--YNAIVTERKNGDYSSFAVTVGTKIP 67
Db 126 AGLIIAVESTTATQPPATPTSDAPPAEQDDPIELVVTGEQDRYVPTASTATKTD 185
Qy 68 ASLREIPQSVSITNQVKDRNVDTFDQARKTPGLRLVNSDGRSSVYARGYEYSEVNI 127
Db 186 TPVRDIPGSIQVPRILEQDKTTRIOEVLQNVGVNKGNYGGTD--ACGYRIRGPDQ 242
Qy 128 DGLPAQMSQING-----TLNLFADFVEMRGPSGLFDSGSGMGGINLVLRKPTKA 180
Db 243 DG-----NFRNGFNDDTFYSLVDTANIDRIEVLKGPASVLFQGAEPGGILNVVTKQPLRT 297
Qy 181 FQGHAAAGFTHKQYKAEADVSGSLNSDGSV--RGRVMAQTGASPRPAEKNRHETFYA 238
Db 298 PYAAELNVGNVAFYRPSFDISGLTDDGSLIYRLNVAQNSGSP----RDYNFLERVFVA 354
Qy 239 -AADNDINPDTVLGAGLYQQRHLAPYNGLPADANKLPSPQHPVFGVADMNKFKONSHD 297
Db 355 PVITWNISDRTSLTFBLEYQDNDYLPDRGIPSGIDRPAP-IPISRFVGLP-HVYNDSTFR 412
Qy 298 VFADLKHPGNGGKVGKVRYSRDRDADSNYAFAGSLGKMKITAGPCGNTADDKACAVGL 357
Db 413 IGYRLEHDFSKDWQLRNAPSFVSGKSSGTAYGGVDL-----IDDQFAPITV 459

Qy 358 G-TETKOKALAFDASYSRPRLNGTANEFVIGADYNFRSTNEQGRITTLTARGGLALNEF 416
Db 460 SRDEFTRDIVTLQTEVVGVKFTGSIHQPLIGVELRR-----NTWKYTSFDVADPILLDIF 515
Qy 417 RSIPQVDLIANARKGVRSYHTVATENLDEFGIYKSTFHPADGLSLIGGRLGHY-KIE 475
Db 516 NPNVDVLPATPDSESTFSY-----TTRRDITLGIYVQDQITFADNLKLLVGGRFDAFORKE 570
Qy 476 SGEKTLHKASKTKFTGYAGAVYDLNNDNSLYLSLYTPQT--NLDAAGKLLKPRQGN 533
Db 571 EGPSETASEESLSAFSPRIGIVQPIQAIISLYASYSSQSFKEDRFGRSASNEPFKPTRGT 630
Qy 534 QFEVGYKGSYNDRLNARVSPYRMKDKNAAA--PLNPNKTRVAAALGKRVMEGVETEIS 591
Db 631 QYEVGIRKAD-ISEKLSATLAAEYITKTNVTSNDPNDLS-----VQVGEQSRSGIELDIG 685
Qy 592 GAVTPKQIHAQYSLHSQIKTASNSRDDGIFL-----LMPKHSANLWTTVQVTPELTIG 646
Db 686 GEIVPGWNIITASYTY-----TDAITSKDNITPVGNRIDNVNPEHAASLWTSYEL----- 733
Qy 647 GGVNAMSGITSSAGMHAGYATFDMAAAYRFTPKL-----KLQINADN 689
Db 734 -----QSGDLKGLGFLGYVGDYADVENTSLSSYFRTDSAIYKRWNRWRLALNFRN 788
Qy 690 IFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
Db 789 LFNETYVETSQARNTI-YPGAPFTVIGSFQIF 820
RESULT 17
AB1944
ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB1944
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-885 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073058.1; PID:gl7130447; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: al1101
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.9%; Score 486.5; DB 2; Length 885;
Best Local Similarity 23.9%; Pred. No. 7.9e-26;
Matches 179; Conservative 117; Mismatches 343; Indels 109; Gaps 26;
Qy 7 VFRINMTAATYLAALSSSVFAAQTAADLETVHIKQRSYNAIVTERKNGDYSSFAVTVGTGI 66
Db 191 IFEVAAPTAPTAPQESQVQPEQS-----ASEEPIELVVTGEQGYKVEDATGGTGT 244
Qy 67 PASUREIPQSVSITNQVKDRNVDTFDQARKTPGLRLVNSDGRSSVYARGYEYSEYN 126
Db 245 NTPIRDTPFSIQVVPVEEVKQQVQVTDALRAVPG--VISQDAPVSA-----FESFN 295
Qy 127 IDGLPAQMSQINGTLPNLFA-----FDRVEVMRGPSGLFDSGSGMGGINLVLRKPT 178
Db 296 PRGSSGFLNGRIRDETIGTAGSGVANIERIEVRGPAGALFGSGPGGVNIVTKQPL 355
Qy 179 KAFQGHAAAGFTHKQYKAEADVSGSLNSDGSVGRVMAQTGASPRPAEK-----N 230
Db 356 SSPFYIEGTGVSFDIYEGRVDLTGPLNRDND---TLLYRLTAS---ASKLGGFIDFVD 408
Qy 231 NRHETFYAAADW---DINPDTVLGAGLYQQRHLAPYNGLPADANKLPSPQHPVFGVADW 288


```

Db 409 NERYISVLTWLDIKNNLTLEAEYLSKTN--PYNNGLPA-LGTVLNPBEGEILPSRNL 465
Qy 289 NK--FRNNSHDVFA--DLKHVFGNGGYGKVGMRYSDBRDAOBNVAFASGKLGMKTPAGRP 343
Db 466 NEPSDFKDKDROVYRRLGVYNEHFRFSENNWOFRNLSRAAFOEYQNEALFPELTL----- 516
Qy 344 GCNTADDDKACAVGLETEIK--QKALAFDASYSRPFRLGNTANEPIGADYNFRSTNEOG 401
Db 517 ----EDRRTVLRFQASLESTBNVYLVLTNVVGDKRTGSIARKLIFGDLTDDTRNDT 571
Qy 402 RTTLVYRAGGLALNFRSRIPOVDLIANARKGVGSHVTAENLDEFGYKGSFTHPADGL 461
Db 572 QDFAYE-----PINLNPVYGTOITIGSSAPSSNFINT-----DLGILYLOQMTLADNF 621
Qy 462 SLIGGGRUG--HYKIESGEGKTLHKASKTKFTGYAGAVYLDNDNNSLYLSQLYTPQTN 519
Db 622 KILLOGREDPDISOKOEDNGEFTINFLQDEAFSPRGLVYOPTKNISLVGSYSRSFTQVVG 681
Qy 520 LDADOKLKLPRGNGFVGYKGSVYMDDLINARVSYRMKDKMAAA--PLNPNNKTRRAA 577
Db 682 TSFDORLQPBEGTEYGVKIGKDWLDRKLTTLAAYLOITRSVLSNDPNPG---FLIQ 737
Qy 578 LGRKRWEGETEISGAVTPKQIHHAGYSYLHSQITKASNRSDGJIFL-LMPGHSANLMTT 636
Db 738 TGEQRSQGIELUDIABEILPGWKILGSIYAY--SDAKYTQDQEFEGULNLTTPHNASLMTT 795
Qy 637 YQVTELTITGGGVNANMSGITSSAGMHAG-----YATFPMAAAYRFTPTLK 682
Db 796 Y----ELQSGN----LQGLGFLGLYVIGEROGDLSNPELPSYRTPDAAMFYR-RNNFR 846
Qy 683 LQINADNLFNRRYVARVGANTFINPGS 710
Db 847 AALNTNNLFDTEYFETAPDALSV-VPGA 873

```

RESULT 18
AE2082
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2082
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuguchi, S.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. Nucleic Acids Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2082
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-854 <KUP>
A:Cross-references: GB:BA000019; PIDN:BA873910.1; PID:G17J31302; GSPDB:GN00179
C:Genetic source: experimental source: strain PCC 7120
C:Gene: a172211
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homology domain 1
Query Match 12.8%; Score 483.5; DB 2; Length 854;
Best Local Similarity 24.1%; Pred. No. 1.2e-25;
Matches 183; Conservative 134; Mismatches 328; Indels 113; Gaps 29;
0Y 18 LAALSSVFAAQTADLETVH-----IKGRSNAIVTEKRGDYSSFAVTGTRKIPASL 70
|||:::|||||:::|||||:::|||||:::|||||
Db 157 LAPINASTIPAPRQPPQPSSETPPAQPTAENNEPIELVYTGNDQGRVRESSATATIDAPL 216
|||:::|||||:::|||||:::|||||:::|||||
0Y 71 REIPQSVSLITMQQVKKDRNVDTFDQARKTPGLRVLNSDDGRSS--VYARGYEYENID 128
|||:::|||||:::|||||:::|||||:::|||||
Db 217 RDIPIASVQVPIPEKIIIOIDRQVAVRLNELANNVSGVQOOSGGYGLSSSGYPIFRG---ESGFE 273
|||:::|||||:::|||||:::|||||:::|||||
0Y 129 GLPAGMOSINGTLPLVLPFA-PDRVEVWRGP-SGLPDSGEMGGIIVLVKRRPKAFQGHAA 186
|||:::|||||:::|||||:::|||||:::|||||
Db 274 GLRNGFKQPGFSTPDDVAIVAEIKVEVFLKGPASVLYGSANPNPGVAVNTITKKPLPDPSYRVG 333
|||:::|||||:::|||||:::|||||:::|||||

```

0Y 187 AGFGHKOYKAEADVSSLSNDGSVGRV--MAQVVGASPPPAEKNNHFFVAA--ADWD 243
Db 334 MTGSSYDFRRTIDFTGTHLDDKSVLYRLNVAIYENSSGSPFDIEN--ESPFSIVTVN 390
0Y 244 INPDVLGAGVLYOORHLAPYNGLLPADANNKLPSLPOHVVG--ADNMKFKONSHVPAJL 302
Db 391 ISPKTSMVFEEYEQKYNVTFDRLGL--PGMTFPQIPISRFLEGGCFNNAEPIISVFPYNL 448
0Y 303 KHYFGNG----GYKGYGMRYSDRDADSNYAFAGSKU--GMKTPAGRCGNTADDAACA 354
Db 449 EHOQSDMKKFRGQGNVVISIRGNTRIA--RNTNFSSEPLEDDGQTLPR--TSETSDBOEN 504
0Y 355 VGLGTEIKQKLAAPDASYSRPFRLGNTANEVIGADNVADNRSTNUGORTLLVYARGALAN 414
Db 505 ISLQTEVSGK-----FNTGSIKHVNLVGEIAKTKFT-----Y 537
0Y 415 EFRSIP--QVDL--IANARKGV--RGYSHTVAENLDEGIYKSTFHPADGLSLIGG 467
Db 538 DFFSAPIASIDIFNPVYGAQGTFRDFRAGYGGDNL--AVYFQNLIEFTPNKLILAGG 594
0Y 468 R---LGHYKIESGEGKTLHKASKTFTGYAGAYVDLNDNNSLYISLSQLYTPQ--TNLDAD 523
Db 595 RFDWIDSEDRDPVSNVTNVESEBSNFSRPRVOIYQPTNISTSLYASMTNSFNPOFFGRSRT 654
0Y 524 GKLLKPRQNGNFEUYGYGSYMDDLNARVSFYRKDKDAAAAPLNPNKKIKRYALGRVM 583
Db 655 GESFPEPTESEQFEVGIQOEFPDKRLSATLAVFDITKNNVLTPDDEVNDFS--VOIGOKS 712
0Y 584 EGVETEISGATYPMKOIHAGYSYLHSQIKTAS-----NSRDDGFIILMPKHSANLMTYQ 638
Db 713 RGLIEDIAGELTLPKMKIATAYATIDSSVSKDNDELRNDRLSGV----PENSASLWTTYE 766
0Y 639 VTPELLTGGGVANMSGITSSAGM-----HAGYATFDMAAVYRFTPKIKLQ 684
Db 769 FQK-----GSLAGCGFLGLVYDEREATLENTIKIPSYVTDASIFYR--RDNWRAA 819
0Y 685 INADNIFNRHYARVAGGANTFNIGSERTWTANLRYSF 722
Db 820 INIKNLDFTEYYE---SOSFYLVPAAPLVLTGITSFEP 854

```

```

RESULT 19
AD2421
ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:/Species: Nostoc sp.
A:/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:/Accession: AD2421
R:/Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Iriyuchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yaenda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:/Reference number: AB1807; MUID:21595285; PMID:11759640
A:/Accession: AD2421
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-858 <KTR>
A:/Cross-references: GB:BA000019; PIDN:BAB76623.1; PID:g17134062; GSPDB:GN00179
A:/Experimental source: strain PCC 7120
C:/Genetics:
C:/Gene: al14924
C:/superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.7%; Score 481, DB 2, Length 858;
Best Local Similarity 25.2%; Pred. No.1,8e-25;
Matches 189; Conservative 127; Mismatches 339; Indels 96; Gaps 27;
OY 17 VLAALSSVFAQAADAEIETHIKGQ-----RSNYALTVEKNKGYSSEFATVGTKIPASL 70
:::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 159 LIVAVTSKKVAMQP--ETPAIPEQPAAGDEPIELVWTEGODGYRPVASVGTIRDTPL 216
:::|::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 71 REIFQSVIIINQQVKDRNVDPQLARKTPGLR-VLSNDGRSGSYVARGEYSEYNI-- 127

```


Db 217 RDIPQIQVPRQVLOEQRATLGLDALRNVGVNTRGSGDRADSFTRGFIFGFGNVLN 276
Qy 128 DGLPAQMSINGTLP---NLFAFDRVEVMRGPGLFDSSGEMGGIVNLVRKPTKAFQGH 184
Db 277 NGLP-----DRTLTRDLYNVEREVLKGPASVLYGLGNPGGVNIVTKQPLANPFYN 330
Qy 185 AAAGFGTHQYKAEDVDSGLSDSGVRGRVMAQTVGASPRPAEKNNRHETPYAANDWI 244
Db 331 IEATVGNVLYRGGIDFSGPLNDSKTLRLNLAQNSGSYIDFVGNRSFFFIAPISAA 390
Qy 245 NPDVTLGAGLYQORHLAPYN--GLPADANNKLPs-----LPQHVFGADWNKFKWNHSD 297
Db 391 GKNNTLTFEGEYSQKTDISRTVVVLPV-VGTVLPGDGRRIIPNRTVTEPEGDTQIETR 449
Qy 298 VFADLKHFGNGGYGKVMRYS--DRDADSNYAFAGSKLGM---KTPAGRPGCNTADDDKAC 353
Db 450 LGYRLHRESENNSLRNDFRVTFEHNADNNQAF---LGLDADNETANRSTYSSES--SN 505
Qy 354 AVGLGTEIKOKALAFDASYSRPFRLGNTANFEVIGADYNNRFRSTNEOGRRTLLYARGGLAL 413
Db 506 IYNLTADI-----SGRFFTGSIHQHLLFGVNMRSRFDNF-----NFGIDL 545
Qy 414 NEFRSIPQVDLIANARKGVYSGHVTATN---LDEFGYKSTHPPADGLSLGGRGLG 470
Db 546 ABELTPLDIYNPV--YRQPIVGRIDAVEDNSRLTDTLGIYIQQDKFAENFOLLGGRPD 603
Qy 471 HY-----KIESGEGTKLHKASKTKFTGVAGAVYDLNDNNSLYLSLSQLYTQT 518
Db 604 LFTQKNKFLDNTTELOAGDA-----FTPRGVIVYKPIAISLYASYSQSFNPE 653
Qy 519 NUDADGKLLKPRQNGQFVGKSYMDRLNARVSFYRMKDKNAAAPLNPNKKNTRYA-A 577
Db 654 GRSADGNLFQPERGTQYEVGVKAD--LNDRISSLTSLYRLTRSNLLT--TDPNN--SRFSIQ 709
Qy 578 LGRVMEGVETISGAVTPKQIHAGSYLHSGIKTASNRDDGIFLLMPKHSANLWTTY 637
Db 710 TGEQRSGIEFDIAGEILPGWKIFAGYATDAEIVEDNTYTSNRLTNVPEHSFNWTTY 769
Qy 638 QVTPELTIGGGVN-AMSGITSSAG-----MHAGGYATFDAMAAYRFTPKLQIINADNIF 691
Db 770 EISSGDFRGLGFLGYIGDRAGLDNSFEVPSYLRITDASIFYR-RDLRLPALNNKIP 828
Qy 692 NRHYARVGGANTFNPISGERTWTANLRYSP 722
Db 829 DTDYFVSVTNRD-FVLRGDPFTISGTISWDF 858

RESULT 20
D87252
TonB-dependent receptor [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 14-Sep-2001
C:Accession: D87252
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: D87252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-851 <STO>
A:Cross-references: GB:AE005673; NID:gl3421118; PIDN:AAK22016.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0028
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.7%; Score 479; DB 2; Length 851;
Best Local Similarity 24.7%; Pred. No. 2.5e-25;
Matches 198; Conservative 126; Mismatches 341; Indels 138; Gaps 33;

Qy 10 INMTAATVLAALSSVFAAQADLETVHIKQORSYNALVTEKNGDYSF-AVTV----- 62
Db 95 LSIAAVAGVAGLG---LAQAA-----VAGPSADAATVTANADDSVSVTTDAKRVA 144
Qy 63 ---GTKIPASLUREIPQSVSIITNQOVKDRNVDTFDQLARKTPGLRVLNSDDGR---SSVY 116
Db 145 DPSSSKETAPLDVTPKSVTVIPAKIIETQATSLADILRTSPGITFGAGGEGQPLADRP 204
Qy 117 ARGYEYSEYN--IDGLPAQMOSINTLPLNFAFDRVEVMRGPGLFDSSGEMGGIVNLVR 174
Db 205 IRG-QASANNVFDG---VRDSCGVREIFNLQEVVVKVGPDSAYGGRSGGGGGINLSS 259
Qy 175 KPPTKACQGHAAAGFGTHQYKAEDVDSGLSDSGVRGRVMAQTVGASPRPAEKNNRH 233
Db 260 KSPKADSFARGSVGVGTDAVVRATADLNHALNESAVRLNLLA--TQGDTPGRKSVSFD 317
Qy 234 ETPVAAADWDINP-----DVLGAGLYQORHLAPYNGLPADANNKLP-----SL 278
Db 318 -----WGVAPSLATGLDGTQLTASYHLEGDQTPDYGVPPLTKTTQRTASGLIDV 369
Qy 279 PQHVFGADNMNFKXNNSHDV--FADLKHYPGNGGYGKVMRYS-----DRDADSNYA 328
Db 370 DRRSFYGVASRDYQKTKSDIATFA-IDHIDETLNLQVVRYSKSLNDYIVTNPGDGAA 428
Qy 329 -FAGSKLGMKTPAGRPGCNTADDDKACAVGLGTEI--KOKALAFDASYSRPFRLGNTANEF 385
Db 429 QFVGQGMWKR-----GTTKRWNPETVAAVTDLHGKKTFLGLESHSPDVGLLESGREEN-- 481
Qy 386 VIGADYNRFRSTNEQ-----GRTLLVARGGLALNEFRSIPQVDL---IANARKGVRGY 435
Db 482 -LNATYSTFTTSGAACCTGFTIAATTLASLAGDCTLVYRPNDKDAWTGVNRPAPARNV 540
Qy 436 SHTVATENLDEFGYKSTHPPADGLSLIGGRGLGHYKIESGEGTKTLH----- 483
Db 541 AKTTA-----LYGPDTVKFGKVLNGLRHRDYESKVDVATQANGVFTSVTVTP 592
Qy 484 KASKTKTGY-AGAVYDLNDNNSLYLSLSQLYTP-----QTNLDADGK-----LLKP 529
Db 593 RSGSWAFTNYQVGLVYKPTPGSSLYSVYSTASTPPGISAGDQNSNTATGTGNLATVQLP 652
Qy 530 RQNGQFVGKSYGMDRLNARVSFYRMKDKNAAAPLNPNKKNTRYAALGKRVMEGVETE 589
Db 653 EDSSEFAGAKANVPHDTLALSALFQTSRKNQAIQID----ATTYAQVGEVKGFEFG 708
Qy 590 ISGAVTPKQIHCAGSYLHSG-IKTASNRDDGIFLL-MPKHSANLWTTYQVTPELTIGG 647
Db 709 VSGNITPKQVFGGYTYMDSSELVRGAVTSVNOGDPPLANTPKHSISSFTYKVRKIALGG 768
Qy 648 QV-----NAMSGITSSAG-----MHAGGYATFDAMAAYRFTPKLQIINADNIFNRHYAR 698
Db 769 GAYHVSFKSGNGOGAGGASRIYAPAYWRYDAPASWAVSTGVDLQNLQNLTDERYIAR 828
Qy 699 VGGANTFN-IPGSERTWTANLRY 720
Db 829 TNGVHADPAPGROAILTINVKY 851

RESULT 21
A12077
ferrichrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: A12077
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi
Nakazaki, N.; Shimpo, S.; Sugimoro, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: A12077
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-863 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA073874.1; PID:g17131266; GSPDB:GN00179
A:Experimental source: strain PCC 7120

C:Genetics:
A:Gene: alr2175

C:Superfamily: ferrichrome-iron receptor 1; conB-dependent receptor amino-terminal homol

Query Match 12.6%; Score 475; DB 2; Length 863;
Best Local Similarity 25.5%; Pred. No. 4.9e-25;
Matches 176; Conservative 120; Mismatches 301; Indels 94; Gaps 26;

```
QY 47 IYERKGDVSSFAVTVGTIPASLREIPOSVSITINQOVDRVDFPDQARKTPG-LRY 105
DB 201 VTTGEADSRVPAASATRTDPLRDIPOSIVVPOOVTKDQOVVLTNKLQVNSGYIQT 260
QY 106 LSNDDRSSVYARGYEY-----SEYNIDGLPAQMOSINGTLPLFAFDREVRGSGPLF 160
DB 261 ASVYSQFASFTTRIGFNSPDQGNPFTRNGLYGRFGS-QGT-NFSNIERLEVLRGGSVL 317
QY 161 DSSGEMGGIVLVKRPPTAFQGHAAAGFTHKQYKAADVSGSLNSDGSVGRVMAQTV 220
DB 318 FQSGNPGGTYINITYKQPLSEPPYVEAIGSYDFYRGALDLSGPLDDSKTALYR-----L 372
QY 221 GASPRPAEK--NNRHETFYAA--DWDINPDTVLGAGVLYOQRHLAPYNGLPADANNKL 275
DB 373 NMSYEKADNFVDFNDRKENSVASTLSFALGENTTLTLDQETNKNQGYINGVPA-VGTVL 431
QY 276 PSL-----PQHFVGVADNMKFKMNSHDVFADLKHYPGNGYGVKVMRYS--DRDASNYA 328
DB 432 PNLNGRIPIHNRISIGQADSTYSPEIVAVGVNLEHKFSEDMILNAPFYSHYNNTRDTYFA 491
QY 329 FAGSKLGKMTPRAGPCGNATDDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFGIVG 388
DB 492 NA---LDPQORTLQORVODADDRVQTYDLSITNVGK-----FSTGSIKHOLLFG 537
QY 389 AAYNRFRSTNEOGRTLVYARGGLALNEFRSIPQVD---LIANARKGVYSTVATENU 444
DB 538 VDISRY---DQYQSEYEGRTGTPDLFENDYRPERFEVFA-----GNEATLT--- 582
QY 445 DEFGYKSTFHPADGLSLIGGRGLGHYKIESGE--GKTLHKASKTKFTGYAGAVTDLND 502
DB 583 DSLGVYIDQOVFAENPKLLGSRFPIFEQTKDRLSNTEQFQSGSAFSSRVGIYVQPIR 642
QY 503 NNSLYLSLSQLYTPQTNLDADGKLKPRQGNQPEVGVKGSYMDRLNARVSFRMKDKNA 562
DB 643 PLSLYASYSRSPFTIGRASDGEQFPGRGTYEIGVKAD-INEKLSATLAFDLTRSNV 701
QY 563 AA--PLNPNKKTRVYALGKRVMEGVETELSGAVTPKQIHAGYSYLHSQIKTASNRD 620
DB 702 TTNDPAPNPEPS---IQTEQONSRGIELNVAAGILPGMNIIAGYAVTDARI-TQDNSLPI 756
QY 621 GTFLL-LMPKHSANLMTTYQVTPPELTIGGVNAMSGITSSAGMHAGG----- 665
DB 757 GKRLLNVPFHSLSLMTTY-----ELQKGN-----LQGLGFLGLFYLDRCQDLANSFNLS 808
QY 666 YATFDMAAAYRFTPKLKLDINADNIFNRHY 696
DB 809 YLRDAIFYK-RDRFRALNINRLFDLKYF 838
```

RESULT 22

E64817
probable membrane protein b0805 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002

C:Accession: E64817

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; PMID:97426617; PMID:9278503

C:Accession: E64817

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-760 <BLAT>

A:Cross-references: GB:AE000182; GB:U00096; NID:g1787015; PIDN:AACT3892.1; PID:g1787024,

A:Experimental source: strain K-12, substrain MG1655

C:Superfamily: ferrichrome-iron receptor 1; conB-dependent receptor amino-terminal homol

C:Keywords: transmembrane protein

F:17-33/Domain: transmembrane #status predicted <TM>

Query Match 12.4%; Score 470; DB 2; Length 760;
Best Local Similarity 22.8%; Pred. No. 9e-25;
Matches 170; Conservative 109; Mismatches 331; Indels 134; Gaps 23;

```
QY 65 KIPASIREIPQSVSITINQOVDRVDFDQARKTPGL-RVLSDDGS---SVYARG 119
DB 63 KPSRPVADTTMTVISEQVIRKQCATNLTDLAKVPGAFGAFAGENGSTGDAIMYRG 122
QY 120 YEYS-EYINIDGLPAQMOSINGTLPLFAFDREVRGSGPLDSSGEMGGIVLVKRPRT 178
DB 123 ADTSNSIYIDG---RDIQSVSRDFTFTEQVEVLKPSGTDYGRSAPTSINM2SKOPR 178
QY 179 KAFQGHAAAGFTHKQYKAADVSGSLNSDGSVGRVMAQTVGASPRPAEKNNRHETFYA 238
DB 179 NMSGIDASISGSAMFRGTLVNVQYIGDTPVALRVLMVGEKTHDAGRKVKR----- 232
QY 233 AADWDINPDTVLGAG-----LYOQRHLAPYNGLP-----ADANNKLPS 277
DB 233 ---YGVAPSVARGLGTANRLYLVNLYHTQHNTPDGGIPITGLPGYSAPAGTALNHSKG 289
QY 278 LQHFVGVADNMKFKMNSHDVFADLKHYPGNGYGVKVMRYS--DRDASNYA 334
DB 290 VTHNHYGTDSDYDSDTDTATMRPEHDINDTTIRNTTMRVRYQD--YLMTALMGAS 347
QY 335 GKMTPRAGG---CNTADDKACAVGLGTEIKQKALAFDASYSRPFRLGNTANEFGIVAD 390
DB 348 NITQPSDVNSMTWRTANTK-----DVSNKLLTNTQNLSTFTYTSIGHDVSTGB 399
QY 391 YNRFRSTNEOGRTLYARGGL--ALNEFRSIPQVDLIANARKGVYSTVATENUDEF 447
DB 400 FRETQTN-----YGVVPVTLPAVNIYHPDSIHPGLTRNGANANGOT-----DTF 446
QY 448 GITYKSTFHPADGLSLIGGRGLGHYKIE-----GSEG-----KTLHK 484
DB 447 AIIYAFDTLQITDFELNGGIRLDNHTTEYDSATACGSGSGAITCEPTGAKSPVTTVOT 506
QY 485 ASKTKFTGY-AGAVYDNLNNSLYLSLSQLYTPQ-----TNLPADGKLKPRQ 532
DB 507 AKSGNLMKAKGALHYLTNGVNTINYAVSQPPGANNFALAQSGSAGNANTRDFKQQA 566
QY 533 NOFEVGVKGSYMDRLNARVSFRMKDKNAAPLNPNNKTRVYALGKRVMEGVETELSG 592
DB 567 NTSIGTKMQVLDKRLLLALFRDIEVEY---EQNDGTYSGQKRVGVEISVAG 622
QY 593 AVTPKQIHAGYSYLHSQIKTASNSRBDGIFLL--MPKHSANLMTTYQVTPPELTIGGGYN 650
DB 623 NITPAQVIGGYTQOKATIKNGKDVAAQDSSSLPYTPBEAFTLMSQYQATDIDISVAGAGR 682
QY 651 AMSGITSSAGMHAG-----GYATPDMAAAYRFTPKLKLDINADNIFNRHYAR 698
DB 683 YI-----GSMKSGDGAAGVTPAFTEGIVADAKIGYVRNRLDQNLNLYNLFDDIYAS 736
QY 699 VQGANTFNIPGSERTW--TANLRY 720
DB 737 INKSGYRHYRGPRTPLFLANMHF 760
```

RESULT 23

AC2140
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A>Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AC2140

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saeamoto, S.; Watanabe, A.; Iriuch

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-867 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074373.1; PID:gl7131767; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2674
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
Query Match 12.4%; Score 467.5; DB 2; Length 867;
Best Local Similarity 24.7%; Pred. No. 1.6e-24;
Matches 186; Conservative 127; Mismatches 264; Indels 175; Gaps 38;
Qy 61 TVGTKIPASLRIPQSVSIITNQOVKDRNVDTFDQLARKTPGLRLVLSNDDGRSSVYA--- 117
Db 201 SVGTRTDTPLDVPQAIQVPOEVINDQGRSLGVLKNTS-----SASSGRSTSQAPAL 255
Qy 118 ----RGVEYSEYNIDGLPAQMSINGTLNLFAPORVEMRGPSGLFDSGSGEMGIVNLV 173
Db 256 TPVIRGFESNLLRNGLRDLSRFQSEIANV---ERVEVLKGPASVLFQGGDLGGVNLV 312
Qy 174 RKRP--TKAFQ--CHAAAGFTGHQYKAEADVSGSLNSDGSVRGRVMAQTGVGASPPAEK- 229
Db 313 TKQPLNTPYISIGYQVQFGLH--RPTIDFSGLDKDG-----VAYRLNAAQTAESF 363
Qy 230 --NNRHETPYAADWDI--NPDTVLGAG--YL-YOORHLAPYNGLPA----- 269
Db 364 KDFENSESEFFIAPVRLIGNENTNLTAIEYLKYSFETAP--DLPASGTVISNPNGRVS 421
Qy 270 -DANNKLPSPQHPVFGADWNKFKMNSHDVADLKHFGNGGKVGKMYSDR---DADS 325
Db 422 RETNLGEPSLSE-----SESLVRL-----GY-QLDHLNLDNWTIKSEF 459
Qy 326 NYAFAGSKLGMKTPAGR---PGCNTADDRKACAVGL---GTEIKQ-----KALAFDA 370
Db 460 STAF-----LDVPESTVVLPIANSSTSN---GLNRDGRTLRRFLVFNPSSTTSITFNN 509
Qy 371 SYSRPRRLQNTANEFVIGADYNRFRSTNQGRITLYARGGLALNEFRSIPQVDLI--ANA 428
Db 510 SLLGKFKTGISIEHSLLFGEV-----TRETEDRL-----DFROLRDIDIFNPVYR 555
Qy 429 RKGVRGYSHTVATENLDE--FGIYKSTFHPADGLSLIGGRGLG---HYKTESGEGKTL 482
Db 556 PESVSFAIPFGNTWNTKNSLGIYAQDQISLKNILVLGGRLDFWNQDYEDLLSEESF 615
Qy 483 HKASKTKFTGYAGAVYDLNNDNSLYLSLSQLYTP-----QTNLDADGKLLKPRQG 532
Db 616 ER-NDTVFSRPRGIVYKPSENLSLYASYSRSFTPVVGRTRVLDTNTGITTVEGEPEPERG 674
Qy 533 NQFEVGYKSGMDDRLNARVSFRMKQKAA--PLNPNKKTRVAALGRVMEGVE 587
Db 675 TOYEVGLKANLLGDLSTTLAFYNLERTNVAAGLSEPLS-----QTIQKQSRQSGIE 727
Qy 588 TEISGAVTPKQIWHAGYSYHLSQIKTASNR--DDGIFLLMPKXSNLWTTQVTPTELITG 646
Db 728 LDVAGEILPGWNLTAATYATDSKITEDSRPEFDQRLQNVPRNSFGLWSTY-----ELQAG 783
Qy 647 GGVNAMSQITSSAGMHAGG-----YATFDAMAAAYRTPKLLKQLINADNIFN 692
Db 784 ----SLKGLGFGVFTQGERQGLRLNTFTLPSYLRTDASIFYR-RDKFRAAINIQNLFD 838
Qy 693 RHYIYARVGQANTF--NIPGSERTWTANLRYSF 722
Db 839 ENYIE---GARDIVRVIPGAPFTLTGVSVEF 867
RESULT 24
S74450

ferrichrome-iron receptor 2 - Synchocystis sp. (strain PCC 6803)
N:Alternate names: protein sll1406
C:Species: Synchocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74450
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74450
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-828 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAAL6602.1; PID:gl6516167
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: fhua_2
A:Start codon: GTG
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
C:Keywords: iron transport
E:213-346/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:547-828/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>
Query Match 12.3%; Score 465; DB 2; Length 828;
Best Local Similarity 23.1%; Pred. No. 2.3e-24;
Matches 182; Conservative 119; Mismatches 303; Indels 184; Gaps 30;
Qy 17 VLAALSSVPAA---QTADLETVHIKQORSYNAIVTEKNGDYSSF--AVTVGTIKIPASLR 71
Db 143 VLSUTSATAIPENPESEIEW-----ATQCGQGEASYFVPSASTATGLDTPLL 192
Qy 72 BIPOSVSIITNQOVKDRNVDTFDQLARKTPGLRLVLSNDDGR--SSVYARGYSEVNI--- 127
Db 193 DIPQSIQVVPQVQDRNVTELGPAQTVPQ---VSPAGRGTSVFGPGFLIRFPVNS 249
Qy 128 ---DGLPAQ--MOSINGTLPLNLFADPDRVEMRGPSGLFDSGEMGIVNLVRKRTKAFQG 183
Db 250 IFRDGIPIQSLAPLNTT-----DIEQIEVLKGPSSIVFGAGEPGGSINLSKKPLDEPY 304
Qy 184 HAAAGFTGHQYKAEADVSG-----SLNSDGSVRGRVMAQTGVGASPPA 227
Db 305 NAAVSLGNYNDYRLDVLSCGPLPEAIDTVNRLNVSYETSGSFRDFYIGDLWVSP--- 361
Qy 228 EKNRHETFYAAADWDINPDTVLGAGLYOORHLAPYNGLPADANNKLPSPQHPVFGAD 287
Db 362 -----TLTNIGPDTKLNLYQYTTNRTILDEGIPAP---NIADLPNRFLGER 407
Qy 288 WNKPKMNSHDVADLKHFGNGGKVGKMYSDRSDADSNVAFAGSKLGMKTPAGRPGCNT 347
Db 408 FSKFEQDQYLYGTYFNHDFNENLKLHAMQY-----LAYAPRYA-----PLPDF 452
Qy 348 ADDKACAVGLGTEIKQALAFDASYSRPRFLGNTANEF-----VIGADYNRFRSTN 398
Db 453 FDEDT-----GELNRFYYGGGNYQRFFTNLAELIGEFTYGPVXRVHLFGLLEYRNDTETP 506
Qy 399 EQGRTTLYARGGLALNEFRSIPQVDLIANARKGVRGYSHT---VATE-----NLDEFIY 450
Db 507 EFQFSNTFA-----PINVFNPV-----YINTFPPIAPEFFRDQDQNRFAVY 547
Qy 451 GKSTFHPADGLSLIGGRLGHYKTESGEGKTLHKASKT-----DSATQNRSTQSIITDPREEFNQTDNQLTPRVGII 497
Db 548 LDQWMDLFDNLKLLVGLRY-----DTATQNRSTQSIITDPREEFNQTDNQLTPRVGII 599
Qy 498 YDLNNDNSLYLSLSQLYTPQ--TNLDADGKLLKPRQNGQFEVGYKSGMDDRLNARVSFY 555
Db 600 YQPIPTVSLGYSYTTSPNPSFAASLNADGSTFDPTQGRQFEVGVKAD--ITDKLSVTFSAF 658
Qy 556 RMKDKNAAAPLNPNKKTRVAALGRVMEGVETEISGAVTPKQIWHAGYSYHLSQIKTAS 615
Db 659 DIRKQNVPT-IDPANLLFTIQT-GEQTSRGVELYLGSEILPGWNLTAATYATDSKITEDSRPEFDQRLQNVPRNSFGLWSTY 716

QY 616 NSRDGIEILLMPKSHANLMTTYQVYVPELLTGGGVNAMSGITSSAGMHG----- 664

Db 717 TDIVNTLISNVEPSNFSIMTYEI-----QSGNLOGLQFGLFPYVDQREGDL 764

QY 665 -----GYATPEDAWAAYRFTPKLTQINADNIENRHYYARVGANTENI---PGSEKRW 714

Db 765 DNTFVLPSYIFRTDAI FYR-RENMELQINIENTLFTNYLAE---SNDPDLVYPGAPFTV 820

QY 715 TANLRYSF 722

Db 821 VKIGIVTF 828

RESULT 25
AE2129
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AE2129
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritzin, T.; Sasamoto, S.; Matanabe, A.; Iriquchi, S.; Nakazaki, N.; Shikmpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE2129
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-851 <KUP>
A:Cross-references: GB:BA000019; PIDN:BA074287.1; PID:g17J31681; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2586
;superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match	12.2%	Score 460.5;	DB 2;	Length 851;
Best Local Similarity	23.6%;	Pred. No. 4.9e-24;		
Matches 173;	Conservative 120;	Mismatches 316;	Indels 123;	Gaps 25;

```

0Y 17 VLAASSVFAAQTADLETVAHKGORSNAIYTEKNQDYSFPAVTGKIPASLEIPQS 76
Db 164 VLSAPSPQOTPTPEPEPQPAATTDPQIELVVTQGDSDRVYNAATATKTDTPDLRDIQS 223
0Y 77 VSIIINQOVKRNVTPTPOLARKTEGLRVLSNDDRS--SVARGY-EXSEINIGLPAQ 133
Db 224 IQVIRQVIEQORALQOLNEALRNVGVOP-SNSAGRTDRFVIRFDDPNSTYIRGF-KE 281
0Y 134 MOSINGTLPNLFAFDREVVMRGPGLFSSGEMGIVNLVKKRPYK--APQGHAAAFG 190
Db 282 NTSVARETANI---ERIEVLKGPASVILFGOLEPGEVIVINVTQOPOREPBLIMGLEAGSYG 338
0Y 191 THKOYKAEDVSGSLNSDGSYKRGVMAQF-VGASPRPAKNNRHETFFAADWDINPDT- 248
Db 339 F---FRPTVDRNPSPINDSKITLRVNAVAEISESRDDKESRFFIAPVLAMEIGDNTS 395
0Y 249 -VLGAGYLYQOCHLAPYN-GLPADANNKLPSLQHVFGADWNKFKANSHDVPADLKHYF 306
Db 396 LVFDLEYLKDNR---PFRGLVA-IGDRVAEIPFRRIIGEPEDSKVEDLRICTREFHQF 451
0Y 397 GNGGVGKVGMRSDRDADSNVAFAGSKLGMTPARPGCNTADMDKACAV-----GLGHEI 361
Db 452 NDNMKLJRANFRLVSTQTSFVTEPSSL-----DEATGLSRDFGVDRPT 495
0Y 362 KOKALAFDASYSRPRRLGNTANEFVIGADYNNFRSTNEQRTT-----LYARG 409
Db 496 PDEYAFQTDLIGKEFKTGAIEHELIVIGDFENRQTFVDRRGAAPAPALDIENPVYLTAAP 555
0Y 410 GLALNEPFSIPQVDLIANARKGVRSYHTVAATENDEGIYCKSFHFADGLSLIGGRL 469
Db 556 NIESPFRGL-----FADNLT---GIYIQDOIKAENKILIGRY 593
0Y 470 GHYK---IESGEGKTLHAKSKTKFTGYAGAVVYDLINDNNLSLYLSQLYTPQTNLDADGL 526

```

Db	594	DFPNGSPISYDVGVKFDPISDPAFSPRIGIYQPIPTPLSLVASSSRSPAQFNGIGORDNSR	653
Qy	527	LKPRGNGFVCKYKSSVWDDRLNAAVSFPRMKDKAAAP-----LNPNKKTRRYAALGRV	582
Db	654	IEPBGTOIEBGVAREEPLDGLKILASLAGIQTITKVAATPDDADLNF5-----IPVGSVR	707
Qy	583	MEGVETEISGAVTPKWOIHAQSYSLH5QIKTASNSRDDGIFL-IMPKSHANILMTTYYQVTP	641
Db	708	SRGIFEDFIAGELAKOMNIIASAYATDAKI-TEDNTDNGRNLNRPENSASIMTWTYEL--	764
Qy	642	ELTIGGVNNA5GITS5AGMHAG-----GYAFPDMAA5YRFTPKUL	683
Db	765	-----QSGALQIGLGMGVGLFVVEGERODLSNSTFTVGYTRTDALALYR-RDMNMI	813
Qy	684	QINADNIFNRHAY	695
Db	814	GLNFKNIDPVNY	825

RESULT 26
C90739
hypothetical protein ECs0883 [imported] - *Escherichia coli* (strain O157:H7, substrain R1)
C.Species: *Escherichia coli*
C.Date: 18-Jul-2001 #sequence revision 18-Jul-2001 #text change 24-Aug-2001
C.Accession: C90739
R.Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
Sasagawa, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shitagawa, H.
DNA Res. 8, 11-22, 2001
A.Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic
A.Reference number: A99629; MUID:21156231; PMID:11258796
A.Accession: C90739
A.Status: Preliminary
A.Molecule type: DNA
A.Residues: 1-760 <HAY>.
A.Cross-references: GB:BA000007; PIRN:BA834306.1; PIR:Q13360342; GSFDB:GN00154
A.Experimental source: Strain O157:H7, substrain R1MD 050952
C.Genetics:
A.Gene: ECs0883
C.Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

```

Query Match      12.2%; Score 460; DB 2; Length 760;
Best Local Similarity 22.6%; Pred. No. 4.5e-24;
Matches 168; Conservative 109; Mismatches 333; Indels 134; Gaps 23;

QY      65 KIPASLRREIPQSVSIIITNQVKDRNVDTFDQARTEPL-RVLSNDDGRS----SVARG 119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      63 KFSRPVADTRTRMTYISQVIKDQATNLTDALKNVPSPVGAFFAGENENSTYGAIVRG 122

QY      120 YEYS-EVYINDGLPAQMOSINGTLPELFAFDREVYVWRGSGS.FDSSGEMGIVNLVRKPT 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      123 ADTSISYITDGI---RIGGSYRPDTFTEVEYAIKGSBGSDYGRSATSATGSIIMSKOPR 178

QY      179 KAFQHAAGAFCGTHKQYAEADVSGSLNSDGSVGRVVAQTGVSAPREAKXNRHETFYA 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      179 NDSGIDASASIGSAWFRGRGLDVNQIGDTTAVRLNVMGKETHDAGRKXNER----- 232

QY      229 AADWQINDEPYLGAG-----YLYQQRHLAPYNGLP-----ADANKLPS 277
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      223 ---YGVASPIAFGLGTANRLYLTNYLHTVQHTPQGIPTIGLPGYSABSACTATILNHSKG 289

QY      278 LPOHFVFGADNNKFKPMNSHDVYADLKHFYGNQGVGKVMYRSDDRADSNY---AFAGSKL 334
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      290 VDTNHFYGTDSYDSDSTDTDTATMRPEHDINDNTIIRNTTNRSKYQD--YLMTAIMGAS 347

QY      335 GMRKTPAGPG-----CNTADDKACAVGLGTEIKOKALAFDASYSRPFRLGTANAEFVIGAD 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      348 NITGTSIVNSWTMRGTANTK-----DYSNKILITQNTLSTFTFYASIGHDVSTGE 399

QY      391 YNRFSSTEGGRTTLYANGGL---ALNEFRSITPOVDLIANRKQVGRGSHIVATENIDEF 447
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      400 FTRRQTN-----YGVNPTPLPAVNIYHPDSSSHPGGLTRNGANANGQI-----DTF 446

```


QY 448 GIYKSTFHPADGLSLIGGRLGHYKIE-----SSEG-----KTLHK 484
D 447 AIYAFDTLQITRDFELNGGIRLDNTHTEYDSATACGSGRGAIITCPAGVAKGSPVTTVD 506
QY 485 ASKTKFTGY-AGAVYDLNDNNSLYLSQLYTPQ-----TNLDADGKLLKPRQG 532
D 507 AKSGNLVNWKAGALYHLTENGNNVINYAVSQPPGNNFALAQSGSGNSANRTDFKPKA 566
QY 533 NOFEVGYKSGYMDRLNARVSYRMDKNAAPLNPNKKTRYAALGRVMEGVETEISG 592
D 567 NTSEIGTKWQVLDKELLTAALFRDIENEV-----EQNDDGTYSYQKKRVEGEISVAG 622
QY 593 AVTPKQIHAHAGSYLHSGIKTASNRDDGIFLL--MPKHSANLWTTTYQVTELTIGGVN 650
D 682 NITPAWQVIGGYTQOKATIKNGKDVADQSGSSSLPYTPEHAFTLWSQYQATDDISVGAGAR 682
QY 651 AMSGITSSAGMHAG-----GYATPDMAAAYRFTPKLKLQINADNIFNRHYAR 698
D 693 YI-----GSMHKGSDGAVGTPAFTEGYWVADAKLGYRVNRNLDLQNLVNYLFDTDYVAS 736
QY 699 VGGANTFNIPGSERTW--TANLRY 720
D 737 INKSGYRHPGEPRTFLLTANMHF 760

RESULT 27

E85589
hypothetical protein Z1026 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85589
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85589
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-760 <STO>
A:Cross-references: GB:AE005174; NID:gl2513801; PIDN:AAG55177.1; GSPDB:GN00145; UWGP:Z10
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1026
C:Supernfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 12.2%; Score 460; DB 2; Length 760;
Best Local Similarity 22.6%; Pred. No. 4.5e-24;
Matches 168; Conservative 109; Mismatches 333; Indels 134; Gaps 23;

QY 65 KIPASLREIPQSVSITNQVKDRNVDTFDQLARKTPGL-RVLSNDDGRS-----SVYARG 119
D 63 KPSRPVADTRTMTVISEQVIKQCATNLTDALKNVPGVGAFFAGENGNSITGDAIYMRG 122
QY 120 YEYS-EYINIDGLPAQMSINGTLPLNLPADRVEMVRGSGSLFDSSGEMGGINLVLRKAPT 178
D 123 ADTSNSIYIDGI-----RDIGSVSRDTFTNQVEVIGKSPGTDYGRSAPTGSINMISKOPR 178
QY 179 KAFQGHAAAGFTQKQYKAEADVSGSLNSDGSVGRVMAQTVGASPRPAEKNRHETFYA 238
D 179 NDSGIDASASIGSAWFRRTGLDVNQVIGDTTAVRLNVWGEKTHDAGROKVRKNER----- 232
QY 239 AADWDINPDTVLGAG-----YLYQQRHLAPYNGLP-----ADANNKLPS 277
D 233 ---YGVAPSIAGLGTANLRYLNLHVHTQNTPDGPIGLPGYSAPSAGTATLNHSGK 289
QY 278 LPQHVFGADWNKFKNSHDVFDADLKHVFGNGYGVKVMRYSDRADSNY---AFAGSKL 334
D 290 VDTNHYFGTSDSYDDSTDDTATMRPEHDINDNTTIRNTRWSRVKQD--YLMTAIMGAS 347
QY 335 GMKTPAGRG-----CNTADDKACAVGLGTEIKQKALAFDASVSRPRLGNTANEVIGAD 390
D 348 NITQPTSDVNSWTWERTANTK-----DVSNKILTNQNTLTSTFTYASIGHDVSTGVE 399

QY 391 YNPRSTNEQRTTLIYARGGL---ALNEFRSIPQVDLIANARKVGRVGSHTVATENLDEF 447
D 400 FTRETQTN-----YGVNPTLPAVNIYHPDSSIHPGGLTRNGANANGQT-----DTF 446
QY 448 GIYKSTFHPADGLSLIGGRLGHYKIE-----SSEG-----KTLHK 484
D 447 AIYAFDTLQITRDFELNGGIRLDNTHTEYDSATACGSGRGAIITCPAGVAKGSPVTTVD 506
QY 485 ASKTKFTGY-AGAVYDLNDNNSLYLSQLYTPQ-----TNLDADGKLLKPRQG 532
D 507 AKSGNLVNWKAGALYHLTENGNNVINYAVSQPPGNNFALAQSGSGNSANRTDFKPKA 566
QY 533 NOFEVGYKSGYMDRLNARVSYRMDKNAAPLNPNKKTRYAALGRVMEGVETEISG 592
D 567 NTSEIGTKWQVLDKELLTAALFRDIENEV-----EQNDDGTYSYQKKRVEGEISVAG 622
QY 593 AVTPKQIHAHAGSYLHSGIKTASNRDDGIFLL--MPKHSANLWTTTYQVTELTIGGVN 650
D 623 NITPAWQVIGGYTQOKATIKNGKDVADQSGSSSLPYTPEHAFTLWSQYQATDDISVGAGAR 682
QY 651 AMSGITSSAGMHAG-----GYATPDMAAAYRFTPKLKLQINADNIFNRHYAR 698
D 683 YI-----GSMHKGSDGAVGTPAFTEGYWVADAKLGYRVNRNLDLQNLVNYLFDTDYVAS 736
QY 699 VGGANTFNIPGSERTW--TANLRY 720
D 737 INKSGYRHPGEPRTFLLTANMHF 760

RESULT 28

A83043
probable outer membrane protein PA4837 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83043
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.B.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: A83043
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <STO>
A:Cross-references: GB:AE004896; GB:AE004091; NID:g9951094; PIDN:AAG08222.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4837

Query Match 12.2%; Score 459; DB 2; Length 708;
Best Local Similarity 22.7%; Pred. No. 4.8e-24;
Matches 172; Conservative 135; Mismatches 329; Indels 122; Gaps 25;

QY 22 SSSVFAAQT---ADLE--TVHIKQORSYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQS 76
D 16 SSPALWAETESPALEVLTVTAEEAEGPV-----QYRANRTASATRTDIEDIPQA 70
QY 77 VSIITNQVKDRNVDTFDQLARKTPGLRVLSNDDGRS---SVYARGYSEYNIDGLPAQM 134
D 71 ISVVPQVLDLDSARIERALDFAGGVSQRNNFGLTMFEYNVRGFTTSEYRDFGSANR 130
QY 135 QSINGTLPLNLPADRVEMVRGSGSLFDSSGEMGGINLVLRKRPTKAFQCHAAAGFGTHKQ 194
D 131 GYMNA--PDSATIERVELKGPASSLYCRGDPGCTVNLVTKKPOAERPARLHASAGSWDR 188
QY 195 YKABADYSGSLNSGSGVGRVMAQTVGASPRPAEKNRHETFYA-----AADWDIN 245
D 189 YRSTLDLNTPLDEBGLLYRM-----NLAVEDSKGFRDYADGQRLLVAPSISWOLD 239
QY 246 PDT-----VLGAGYLYQQRHLAPYNGLPADANNKLPSLPQHVFG--ADNKKFQMNSHD 297

Db 240 PPTSLLEAVEVANNRQVFBGTVAPH-----NHLGSLPRSRFFGEPDDGKIINNNET 291

Qy 298 VPADLKHGVEGNGYGVKMGYSRDRDADSNVAFAGSKLGMKTPAGRPGCNTADKACAVGI 357

Db 292 LQATLRHHNEQMSURLASHYKMGHLD-GVASENSSLA-----AGGYSL 334

Qy 358 GTEIKOKALAFDASYSR-----PRLGNTANEFVIGADYNNFRSTNEQGRITLLYARGGLA 412

Db 335 RREYRRDEEMWDSITQDLGLDHTGSIHQMLMGLVEYRHA--NDE-----LI 382

Qy 413 LNEFRSIGPOVDLIA-NARKGVRGY-----SHTVATENLDEFGIYKSTFHPADGL 461

Db 383 L--RSIPRNPYADIRREPVYQKPPRPGDRDNH---EVDAMALNLQDQIEFSEKK 435

Qy 462 SLIGGRLGHY-----KIESGE-GKTLHKASKTKFTGYAGAVYDLDNNNSLYLSQL 513

Db 436 RGLLGRFPRRYRDMNATLNNRFRSTSSQGTQRAATRIGLYQATPEVGLFANAASIS 495

Qy 514 YTPQTNLDADGKLLKPRQGNQFEVGYKGYMDRLNARVSFYRMKKN--AAAPLNPNNK 571

Db 496 FRNGSGTDMAKAFDEEGRGYBAGVYKLDLDGRLGMTLAAFLKKNVLTADPSNPGYQ 555

Qy 572 KIRYALGKRVMEGVETISGATTPKQIHAGYSYLHSQIKTASNSRDDGIFILMRKHA 631

Db 556 QT---AGEARSQGFPLQFSQGLTEQRLIGAYAYIDAETKQDENTARSGRLNVPKHSQ 611

Qy 632 NMTTYQVPELITIGGGVNA-----MSGITSSAGMHAAGVATPDMAAYRFTPKLKIQ 684

Db 612 SLMGYEFREBGMHAGADAGAAVNYVGERACDSSDSGFELPATYTTVDLLARYPLASNAITLG 671

Qy 685 INADINENRHYAVRGANTFNIPGSEERTWTANLRYSF 722

Db 672 VVVNNLFDRRYER-SYNNVWVAPGEPRLTMSLTINY 708

RESULT 29

G97408
fega proteoin U61401 [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C.Species: Agrobacterium tumefaciens
C.Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C.Accession: G97408
R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markel, B.; Science 294, 2323-2328, 2001
A.Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A.Reference number: A97359; PMID:11743194
A.Accession: G97408
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-747 <KUR>
A.Cross-References: GB:AB007669; PIDN:AAK66224.1; PID:g15155326; GSPDB:GN00169
C.Genetics:
A.Gene: AGR_C_718
A.Map position: circular chromosome

Query Match 11.9%; Score 450.5; DB 2; Length 747;
Best Local Similarity 25.0%; Pred. No. 2e-23;
Matches 195; Conservative 104; Mismatches 358; Indels 123; Gaps 33;

Qy 2 GQFMSVFRINMTATVLAALISSVFAAQTADLETVAHKGRSYNAIVTEKNGDYSSFAVT 61

Db 24 GMEFLSLALGTAAVLMGPVMG--HAEFTVLKQITVEGGGAENA--TGPVRGYVAKKSA 79

Qy 62 VGTIKIPASIREIPQSVSIITNOQVKNR--VDTPDQALARTPRGLRV--LSNDDRSSVYAR 118

Db 80 TQSKTETETKALIPQSVSVGRQEMDRGAATKIDEVALRTYPTAEPFGTPTDPMFYIR 139

Qy 119 GYEVSEYNI--DGLPAQWOSINGTLFNLFAFDEVEVMRGSGFLFDSGEGGIVNLVRRK 176

Db 140 GFGATQGTGFLDNLNLFYSGFGFGQMDAYGLERVENVLKGSASTLYGGANCGGIVQWVSKR 199

Qy 177 ----PTKAFQ-----GHAAGF-----GTHKQYKAADVSGSLN-SDGS--VRGR 214

Db 200 AODTPEVRETEIGINNGCAPFGFDLGDKVDABGVK-VFVTKGVSGGNDYTDYSEDLRPF 258

Qy 215 VNAQTVGASPRAPKAKNNRHEFEYAAADMDINPDVLGAGYLYQORHLAPYNGLPADANNK 274

Db 259 IMPQ-ITTEP-DAQISATLYGYFSALD-----QIVHGNFL-----PYGTVVDA--P 302

Qy 275 LBSLPQHFVVG-ADNMKFKMNSHDVFADIKHYFGNGYGVKMGYSR-----320

Db 303 FGLDKRKAIFYGEPDIDNGRVYQSMVGYDVSHFEDNGMKISQARVGHLYKHETGYPYGGW 362

Qy 321 RQADSNVAFAGSKLGMKTPAGRPGCN--TADKACAVIGLSTIKOKALAFDASYSRPF 378

Db 363 ANADAN-----GQPIIDPITNDVMTLRFEGDGLSKVDSFEVDNRIESQFDT 408

Qy 379 GNTANFVIGADYNNRFRSTNEQGRITLLYARGGLALNEFRSIGPOVDLIANARKVGR-----434

Db 409 GAVNHSPFLGLDYKYRLDQVACCSNAILGAL-----KPYVSTOG 450

Qy 435 ---YSHVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGE--GKTLHKASK 487

Db 451 TNFVYADNIVTQ--QQIGIYAQDQALFEGDGLVTLNGRYDYVDTLNNRLPAGVSRSD 508

Qy 488 TKFTYAGAVYDLDNNNSLYLSQLYTPQTNLDADGKLLKPRQGNQFEVQ--YKGSYMD 545

Db 509 DALSGRAGLADEFDNGLTPYVSAATFENPLIDTLADGPASPEBGHQPBAQIKYBBSFPD 568

Qy 546 DRLNARVSFYRMKDKAAAPLNPNNKTRYAALGKRVMEGVETISGATTPKQIHAGYS 605

Db 569 GSITTSV-FKLYKD-NAIVSYTAGVTTIS--GQPGQVESTGFELKAKNLDENKALASIS 625

Qy 606 YLHSQIKTASNSRDDGIF-LMPKHSANLMTTYOVTP-----LTIIGGV--NMSGITSS 658

Db 626 YTDLDITKANPNLIGKSPWIVPAHTASLWVDYAFDFEFEGISIGGVRYGOKSWADA 685

Qy 659 AGMHAGGATPDMAAYRFTPKLKIQINADINFNKHYAVRGANTFNIPGSEERTWTANL 718

Db 686 NTLRYSDAVFPAAIRYE-KNDWMTASVNVANVFDEKRYKSCAGSVCGW-GDSRTITLKL 743

RESULT 30

A13197
TonB-dependent receptor [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plas
C.Species: Agrobacterium tumefaciens
C.Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C.Accession: A13197
R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Moks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McClell,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A.Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A.Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A.Reference number: AB2577; PMID:11743193
A.Accession: A13197
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-819 <KUR>
A.Cross-References: GB:AB006887; PIDN:AAU45999.1; PID:g17743754; GSPDB:GN00188
A.Experimental source: strain C58 (Dupont)
C.Genetics:
A.Gene: Atu5311
A.Genome: plasmid

Query Match 11.9%; Score 450; DB 2; Length 819;
Best Local Similarity 25.5%; Pred. No. 2.5e-23;
Matches 187; Conservative 106; Mismatches 349; Indels 92; Gaps 31;

Qy 33 LETVHKGGRSYNAIVTEKNGDYSSFAVTGTRIPASIREIPQSVSIITNOQVKNRVDNT 92

Db 134 LQITTYGASAYGHV--DGIYASRSAT-GTKTTPLEIVQOSISVITADYKARGAET 189

Qy 93 FQGLARTPGLRVLNSDDRSSVYARGYSEYNI DGLPAQWOSINGT-LPNL-----144

Db 190 IKEAVNYTAGVHV-----CGSSASTRAFDNIE--IRGFAPTLPLDGLTYPYIGDLGGSP 242
Qy 145 ----FAFDRVEMRPGSLFDSSGEMGGINLVNRKRPKAFQGHAAAGFTGHQYKABAD 200
Db 243 QIDPYLELIEVLKGPSSVLYQNYPGGMINMVSKRPDKPNEVVAGTGT-----DGR 296
Qy 201 VGSLSNDSQSVGR--VMAQTGASPRPAEK--NNRHETFYAAADWDINDPVLGAGYL- 255
Db 297 AYGAFFDFSGPVAGNDAFLRLTGAVTRGTNIDYTKDERFMLAPSFALPKPBDTFTTFLS 356
Qy 256 -YQORHLAP-YNGLPADANNK--LPSLPOHVFVGAD--WNKFKMNSHDVFADLKHVFGNG 309
Db 357 HYKONGVDPYQPLPIGVTKAGPGVAINRDFTEBPAYNGVDRQAVLGVYEFKQFDDV 416
Qy 310 GYGKVGMYSDRDADSNYAFAGS--KLGKMTKTPAGPGCNTADDKACAVGLGTEIKOKALA 367
Db 417 WSIHNAKIYISVDDSVRTFSGYVETGGVTDYTKMRNAIDYSS-----NNQVFA 467
Qy 368 FDASYSRPPRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGLALNEFRSIPQVDLIAN 427
Db 468 TDTNLQAEFTSTGEVGHTVIVGTGDKWFRN-DYTGR---YGFQNTPLDVFN--PRYGSYRE 521
Qy 428 ARKGVGYSHTVATENLDFGLYGVKSTFHPADGLSLIGGRLGHYKIESGE---GKTLHK 484
Db 522 PTVGARW-----DNRIQLGUYAQOI-KWDNWLTLGGRY-DNALQTDNDILSTIGK 573
Qy 485 ASKTKFTGAGAVYDLNDNNLSLSQLYTPQTNLADGKLLKPRQGNQFVGVKGYSM 544
Db 574 KDTAFTAGRLIYLPDNGFAPVSVYSTSPFSGDQGNPFKFTTGEQWEGVLKYPBV 633
Qy 545 DRLMARVSYFMKDKNAAPLNPNKKTRY--AALGKRVMEGVETESGAVTPKQWIHA 602
Db 634 GYDALITVSADFLOKQNV-----PTVDEPTLPAQTGEITHVQGIIEGKATVDFSLDLIA 688
Qy 603 GYSYLHSQIKTASNRDDGIFLLMPKHSANLWTTQV---TPELTIGGV-NAMSGITS 657
Db 689 AASYTDSVYSKADDTQGNKVRFPVNVSLMGKRIEDGPLEGLGFGAGVHRHSSGGVD 748
Qy 658 SA-GMHAGGYATFDAMAAYRF---TPK---LKLQINADNIFNRHYARYVGGANTFN--IP 708
Db 749 AANSFKYPATVVDAAISYDFGKQNPKEGLELNVTAQNLFPDKTY--VSGCSNINSCFY 805
Qy 709 GSERTWTANLRSF 722
Db 806 GKSRAVYANLSYKW 819

RESULT 31
AC2134
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2134
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2134
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-872 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074325.1; PID:g17131719; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: air2626
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 11.9%; Score 449.5; DB 2; Length 872;
Best Local Similarity 24.2%; Pred. No. 3e-23;
Matches 179; Conservative 120; Mismatches 303; Indels 139; Gaps 27;

Qy 47 IVTEKNGDYSFAVTGKIPASLREIPQSVSIITNOQVKDRNV----- 90
Db 206 LVTGEQNRVYRPNASTGRTDITLIRIPQTIQVVPQVQIDQVRTRLRDALLNTGGVVQD 265
Qy 91 ----DTFDQLARKTPGLRVLSSDDG-RSSVYARGYSEYNIIDGLPAQMOSINGTLPNLF 145
Db 266 GGFSTSDOI-----GIRGFGDGTFGSILVDGFK-----DGRGIRETAN----- 307
Qy 146 AFDREVMRPGSLFDSSGEMGGINLVNRKRPKAFQGHAAAGFTGHQYKAEADVSGSL 205
Db 308 -VERIEVLKGPASVLYGQVQGGVINLVTKQPLRDPYNAELSVGSFSTPRPSIDISGPL 366
Qy 206 NSDSVGRVMAQTVGASPRPAEKNNRHETFYAAAD-----WDINPDT-- 248
Db 367 NSDKTLLYRL-----NSVYETSDGFRDNQDVQREFFISPTLKWEIGKATNL 412
Qy 249 VLGAGLYLQORHLAPYN-GLPADANNKLPQLPQHVFGADWNKFKMNSHDVFADLKHVFG 307
Db 413 TLQFDYLANDER---PFRGFLAFGEIIDT-PLERFFGEPDDVRKVEIGLSYLEHNFN 468
Qy 308 NGYGKVGMYSDRDADSNYAFAGSKLGMKTPAGPGCNTADDKACAVGLGTEIKOKALA 367
Db 469 DNWKIRNAFRYQSSDT-FDYRAEPVRLNETTGILTRNFRSNDVYEENVTLTQDVVGK--- 524
Qy 368 FDASYSRPPRLGNTANEFVIGADYNRFRSTNEQRTTLYARGGL---ALNEFRSI----- 419
Db 525 -----FITGSINHLLFLFGVDLAREFTSGTQSRLP-----GVITPSINVFNVPYNAIP 571
Qy 420 -POVDLIANARKGVGYSHTVATENLDE---FGIYGVKSTFHPADGLSLIGGRLGHYKIE 475
Db 572 RPDDELITN-----VVRNQDRSGTGLGIFLQNTAFADNLKLLVGGREFDTVDQN 620
Qy 476 SGEKTLHKASK--TKFTGYAGAVYDLNDNNLSLSQLYTPQTNLADGKLLKPRQGN 533
Db 621 STDLRDGSSEGRYDFAFTPRLGIVYQPIEPISLVASYQSQFPNFGTRVDGSILEAERG 680
Qy 534 QFEVGYKGSYMDRLNARVSYFMKDKNAAP--LNPNNKTRYAALGKRVMEGVETEIS 591
Db 681 QYEVGVKGEFLDGLAATLAAHYHTKSNIAATDLNDPD---FLPLPIGEORNOGIELNVA 736
Qy 592 GAVTPKQIHHAGYSLHSQIKTASNRDDGIFLLMPKHSANLWTTQV---TPELTIGG 647
Db 737 GEISFGNNVTASYSHIDAETRDNDGLOGNRPANVPNTASFTWTTYELOQDGLQGLFGL 796
Qy 648 G---VNAMSGITSSAGMHAGGYATFDAMAAYRFTPKLQINADNIFNRHYARYVGGANT 704
Db 797 GLFVVGDRQGDSSNTYI-IPGLRTDA-AIYKRDNRNAGINIQNLNFKHYL---GANF 851
Qy 705 FNI---FGSERTWTANLRSF 722
Db 852 GRVAIEFGAPLTVIGSPSVTF 872

RESULT 32
AH2626
ferrichrome iron receptor Atu0409 [imported] - Agrobacterium tumefaciens (strain C58, DuJ)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH2626
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2626
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-708 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41430.1; PID:g17738752; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)

C:Genetics:
A:Gene: Act0409
A:Map position: circular chromosome

Query Match 11.7%; Score 441.5; DB 2; Length 708;
Best Local Similarity 25.3%; Pred. No. 8e-23;
Matches 191; Conservative 99; Mismatches 344; Indels 121; Gaps 32;

```
QY 27 AACGADLETVHIGKGRSYNAIVTEKNGDYSFVAVTGTRIKPISRLIPQSVSIITNQVK 86
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 8 AGCTTVLKQITVEGQGENA--TGPRGVYAKKSAAGSKTETETKALIPQSVSVVGGQEMD 65
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 87 DRR-VTFPOLAKRTPLRLV--LSNDGRSSVYARGVSEYNI--DGLPAQMSINGTL 141
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 66 DRRAVTKIIEVLRITPTGVAEPFPTDPTDPTDPTDPTDPTDPTDPTDPTDPTDPTDPTD 125
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 142 PNLFAFDVRYVMKPSGLPFSSGEMGIIVLVKRR-----PTKAFQ-----GHAAGF-- 189
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 126 MDAYGLERIEVLRIPASVLYGKANPGGIIVOMVSKRAODTPVRETETGINNFGNAFFGFDL 185
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 190 -----GTHKQYKABADYSGSLN-SDGS--VRGRVAAQTGASPPRAEKNRHEIFYAA 239
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 186 GDKVDAEGVWK-YRVTGKVSQGNITDYSIDLGFIMPQ-ITFEP-DAQTSATLYGFFSA 242
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 240 ADMNDIPDTVLGAGTYIQQRHLPYNGLPADANNKLPSPFQHVFG-ADWNKFRMNSHDV 298
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 243 LD-----QVHVNGGFL-----PYGTVVDA--PFGKLDKRAFYGEPDIDNGRVYQSMV 288
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 299 PADLKHYFGNGGKGVKGMRYSD-----RDADSNVAFASGKLGKMTPARPGC 345
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 289 GIDVSHFEDGKWKISQNAKRGHLYKHETGPYPCGMANADAN-----GQPIIL 334
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 346 N--TADDKACAVGLGTEIKOKALAFDASYSRPERLNTANEFVIGADYNRFRSTNEQRT 403
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 335 DPTINDYMLTRFGYDGLSKVDGSGVNRLEGQDTGAVNHSPLFGIDYKRYRLDQVQAC 394
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 404 TLYARCGALNERSIPQVDLIANARKVGR-----YSHTVATENLDEFGIYKSTF 455
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 395 GSNALGAL-----KPYVGYSTGTFVYVADNIVIQ--QOIGIYADQL 434
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 456 HPADGLSLGGGRLGHYKIESGR--GKTLAKSKTKFTGYAGAVVDLNDNLSLYLSLQ 512
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 435 RRGDGLVTLNGYVDVDTLNNRLPAGVSRSDNALSGRAGLAIEFNDGLTPYVSAN 494
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 513 LYPQTNLDADGLKLPKROGNOFEVG--YKGSYMDRLNARVSFYEMKDKNAAPLNPNN 570
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 495 FNPPLDITLADGTPAPBEGHOFBAGIKYEPSPFDSITNSV-FKLVKD-NALVSTTAGG 552
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 571 KTRVYALGKRVMEGVETISGAVTPKMOIHAGYSYLHSQIKTASNSRDLGF-LMPKH 629
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 553 VTTT--GQFQVESTGFELEKAKNLDENWKALASYSTDLDTKDANPNLIGKSPWIVPAH 611
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 630 SANLWTTQVTPR---LITGGCV--NANSGITSSAGMHAGVATPDMAAAYFTTKKL 683
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 612 TSLWLDVYFTBETFEGLSIGGVRYQKSWADANLTVRSDAVDAIRYE-KNDWTA 670
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 684 QINADNIFNRHYARVAGANTFNIPEGSERTWTANL 718
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 671 SYNVANVFDEKEYKSCAGVSVCGM-GDSRTITIKL 704
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 33

DB3081
Probable outer membrane receptor for iron transport PA4514 [imported] - Pseudomonas aer
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence revision 15-Sep-2000 #text_change 14-Sep-2001
C:Accession: DB3081
R:Stover, C.K.; Pham, X.Q.; Brwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Laidig, K.; Lam,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; WUID:20437337; PMID:10984043

A:Accession: DB3081
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-753 <STO>
A:Cross-references: GB:AE004865; GB:AE004091; NID:g9950752; PIDN:AA07902.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4514
C:Superfamily: ferrichrome-iron receptor 1, tonB-dependent receptor amino-terminal homol

Query Match 11.7%; Score 441.5; DB 2; Length 753;
Best Local Similarity 22.9%; Pred. No. 8e-23;
Matches 186; Conservative 129; Mismatches 318; Indels 179; Gaps 37;

```
QY 15 ATVALASSVFAQTADLETHIKGRSYN-----AITYEKNGDYSFVAV--TVG 63
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 17 ASAGVATIRAIAPPAQADEA---GQKTDKDRVLSLDAITVEGQDETTYNVRSAS 73
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 64 TKIPASLREIPQSVSIITNQVKNRNVDTFDLARKTPGLRVLSNDGRSS--VYARGV 120
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 74 KKTATLPLTPKTVTVIPQOVIKDTGALFLADALRTTPGITGAGGCGNPAGDRPIRGF 133
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 121 E-YSEVNIIGLPAQMSINGTLNPLFAFDRVYMRGSGLPSSGEMGIIVLVKRRPTK 179
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 134 NMSDTFLDG---MRDVASQTRVFNVEQIEVSKGPGSAYTGAGSTGSLNLSK--TA 187
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 180 AFGCHAAAGF--GTHKQYKABADYSGSLNSDGSVRGRVAAQTGASPPRAEKNRHEIFY 237
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 188 KQDNFTDAGFTMGSDQTRRTTLDVNRIMIDNAAFLNLN-----KIDAIY 232
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 AAAD-----WDINDPTVLG-----AGLYQQRHLPYNGLPADANNKL--PSLPQ 280
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 233 AGRDEVSVSRMGVAPVTGFTPTPATLSYHLSITDMDPDGLPLTVNRSKANSKPA 292
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 281 HV-----FVG--ADWNKFRMNS--HD-----VFADLKHYGN--GGYG 312
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 293 SVDRNFPYGLKORDYKSTTDSGTFRIEHDLNDNLTSNSTRVTRTLDIYVSNPDSRG 352
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 313 KYGMRYSDDDADSNVAFASGKLGKMTPARPGCNTADDKA-----CAVGLGTEIK 362
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 353 NYANGVYVSAS-----RNSTSKGVNQTDLKANETGRTIKTLYTGL----- 396
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 363 QKALAFDASYSRPFRL-----GNTAN--EFVIGADYNRFRSTNEQRTLYARGLALNE 415
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 397 --EFSYEDVHNRPYAITSGGAGNTCNALLASGDOTSLNRPPTPGDNTGSLTDGLAYTD 454
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 416 FNSIPQVDLIANARK-----GVRYGSHTVATENLDEFGIYKSTHPPADGLSIG 466
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 455 TDTKTSAAVVPDTLKLSEQWEINLGLR-----YDDPDT--KSSGYQTAGR-- 498
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 467 GRLGHYKIESGGKTLHKASKTKFTGY-AGAVYDLDNNSLYLSLQVTP-----Q 517
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 499 GPAGYFKRE-----NSHFVNYQTGLVYKPAKPNOSITYLASTSNPFGETGEGQ 548
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 518 TNLDAADGLKLPKROGNOFEVGYKGSYMDRLNARVSFYEMKDKNAAPLNPNNKTRVYA 577
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 549 AIDSVGNNDLDERNRNLELGTKMAFFDDALSLNALPR-TKTNARVASP-DVSLQVLA 606
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 578 LGRVMEGVETISGAVTPKMOIHAGYSYLHSQIKTASNSRDLGFILMPKHSAN--LW 634
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 607 DGEORVQGVLEFGNKLTEKMKVFGGYTLDBEIRKSTVKSDEG--NKMPQTAQNNTILM 664
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 635 TTYQVTPBELTIGG---VNANSGITSSAGMHAGVATPDMAAAYFTPKLQIINADNIF 691
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 665 TTYDLDQNTFTIGGTYTYDKQYGNANS--TYLPSTWRDYAMASIKVSKVNDQLQNLNTL 723
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 692 NRHYARVAGANTFNI-PEGSERTWTANLRYSF 722
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 724 DKRYFPQVYSTMAHVAAPG--RTALLGVNHFH 753
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
```

RESULT 34

A63481

Db	659	VTSVDNGKLDAYVHNNMMLGYQATRDLDKQLQNLNLFDPKAYVERV	7033
RESULT	35		
AE2085			
ferrichrome-iron receptor [imported] - Nostoc sp. (strain P			
C;Species: Nostoc sp.			
A;Name: Nostoc sp. strain PCC 7120 is a synonym of Anabaena			
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_ch			
C;Accession: AE2085			
R;kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamot			
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yama			
DNA Res. 8, 205-213, 2001			
A;Title: Complete Genomic Sequence of the Filamentous Nitro			
A;Reference number: AB1807; MUID:21595285; PMID:11759840			
A;Accession: AE2085			
A;Status: preliminary			
A;Molecule type: DNA			
A;Residues: 1-858 <KUR>			
A;Cross-references: GB:BA000019; PIDN:BA073935.1; PID:g17113			
A;Experimental source: strain PCC 7120			
C;Genetics:			
A;Gene: ali2236			
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent			
Query Match 11.2%; Score 424; DB 2; Length:			
Best Local Similarity 23.8%; Pred. No. 1.8e-21;			
Matches 170; Conservative 124; Mismatches 328; Inde			
Qy	47	IVTEKNGDYSFAVTGKIPASLREIPQSYSIITNQOVDNRNVDTFDQ	
Db	199	VVTGEQDTRYREATTATKLDVLRDPVAPSVQVIPKEVIDRQVIRLNE	
Qy	107	SNDGRGS--VYARGEYS--BYNIDGL-----PAQMOSINGTLPLNLF	
Db	259	ETYGGLASQGYFIRGFSTGETFLRDFGRDFGIFSPRDVANI-----	
Qy	157	-SGLFDSSGEMGGIVNLVRKRPTKAFQGHAAAGFTGHQYKAEADVSSG	
Db	309	ASVLYGATSPGGVVNITKKPLADPPFYQINGTIGNDFYKSSIDFSGP	
Qy	216	MAQTVGA-SPRAPENRHRHETFYA-AADWDINPDTVLGAGLYQQRHLA	
Db	369	NAAYENARSPRDFVEN--ESTFIPIITVKAGERTNLTFGEYOKYDVT	
Qy	274	KLP-SLPOHFVVG-ADWNKFKMNSHDVADLKHFYGGNGYQKGMRYSD	
Db	425	KVPFDLPINRFLGPNLNRGEFKNNFTYTLESEFYGNNNNKKFQAENV	
Qy	332	SKLGMKTPAGRPGCNTADDKACAVGL-GTEIKKALAFDASYSRPFRLG	
Db	480	-----TKGAQPDINADGQTVARGYRNVDEAQNLSFONEISKFNFTG	
Qy	391	YNRPRSTNEQRTTLYARGGLA----LNEFRSI PQVDLIANARKGVRCY	
Db	534	LSYKYKSYD-----FLSGSIDGLDIFNPVYGAPAPDLSRNEEYGG-	
Qy	447	FGYVGKSTFHPADGLSLIGGRLG-----HYKISGEGKTLHKASKTFP	
Db	578	IAYVFQNLVELTPQIKLLAGGRFDVDSVYRDVET--GTVVDETSDSKFX	
Qy	502	DNNSLYLSLSQLYTPQT-NLDDADGKLKPRGNGFEVGYKGSYMDRLN	
Db	636	DSTSIYASWNSFNFPQIFGKTRNNEPKPTAEQFEVGIKQEFLNKRUS	
Qy	561	NAAAPLNPNNKKTRYA-ALGRVMMEGVETEISGAVTPKWQIHAGYSYLH	
Db	696	NV-----LITDPEDDNFSICTGEQKSRGVELDVVGELPGWKLIIGTYTD	
Qy	618	---RDDGIFLLMPKHSANLWTTYQV----TPELTI GGVNAMSGITSSA	
Db	753	VNNRLVGV----PYNASLWITYBELQGNLOGLGLGVVYVVGREASL	

QY 669 FDMAAAYRTPKLTQINADNIFNRHYARVGANTNIGSEKRTWTANLRYSF 722
 Db 809 TDSIFPK-RDNRAALNFKLPTKYE--SOSFIVPAAPFTVIGVSFEF 858

RESULT 36

AC0547
 ferritin gene B receptor precursor [imported] - Salmonella enterica subsp. enterica serov. C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
 C:Accession: AC0547
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Comerford, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov. A:Reference number: AB0502; PMID:11677608
 A:Accession: AC0547
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-696 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08819.1; PID:g16501633; GSPDB:GN00176
 C:Genetics: A:Gene: STY0396

Query Match 11.2%; Score 421.5; DB 2; Length 696;
 Best Local Similarity 23.1%; Pred. No. 2e-21;
 Matches 175; Conservative 121; Mismatches 53; Indels 109; Gaps 25;

QY 12 MTAAATLAA-----SSVFAAQTADLETVHIKGRSYNAIVTEKNGDYSSFAVTG 63
 Db 1 MFATRRALLIGALIGATFPLFAQETTKNDTV-----IVT-----SPVSGA 43
 QY 64 TKPASPAREIPQSVSITTNQVKDRNVDTQOLARKTPGLRVLSNDDGRSS---VYARG 119
 Db 44 TRLATPDIETPOSVSITTRQOFEGATSVRAQVSTPG--VYSNQIGANRPDIYVLRG 101
 QY 120 YERSEVN---IDGLP--AQWQINGTLNLFAPDRVEMVMSGSGSLPSSSEMGIYVLR 174
 Db 102 FSGSGLDNVTYLDGKMGMDTNSHSLVADPWPLEDIEVVGSPASVLYGRSSPGIYSLTS 161
 QY 175 KRPTKAFQGHAAAGFTGHKOYKAADVSGLSNDSYRGVMAQTVGASPRPAKNNRHE 234
 Db 162 RKAFPAAGGVKLPAGNNORGAAPVDTGPDLDNERAAALSGMTRTADSQFPLKEERY 221
 QY 235 TFPAAADWDINPVTVG-AGLYVQ-----QRHLAPYNGLPADANNKLPSPQH 281
 Db 222 ALMPSLTWRITDRTRLDLMAVYLHRDPQGGSHGLPYQGTAVVYNG-----GKISNT 272
 QY 282 VPFGA-DMMKFKMNSHDVPAADLKHVFNNGYGVKVMYSPRDADSNAPAGSKLGMKTPA 340
 Db 273 FFGEDDYDYDRRENNVGNIEHLFENGWSVQKLAHYLTQYTLAQVYAAGWLN----- 327
 QY 341 GRGCNTADDKACAVGLGTEIKOKALAFDASYSRPRFNGTANFVGAYNFRSTNEQ 400
 Db 328 -----ETALNRGYS--GSGEKMSALMDQDQDSVDTGALNRLVLDIYQ-----DRS 374
 QY 401 GRTTLYARGLALNFRSIPQVDLIANARKGVGSHTVATE--NLDFGIYKGSFHPA 458
 Db 375 NHTTGY-----YGAFFPIDAFNVPVGAOPDYITLYGREKHKLRQGTUYLQDQW-SW 424
 QY 459 DGLSLGGGLGHYKTESGGKTLHKA---SKTKFTYAGAVYDLNDNNSLYLSLSQLY 514
 Db 425 DRRRFTLGGYDVSNSND-KLHDSRSLDKNNVSTRALLYLFPNGVAPYLSYSTAF 482
 QY 515 TPQTNLDADKLLKPRQGNFVGYKGYMDRLNARVSPYRKMDKKAAPLPLNNKTR 574
 Db 483 TTSFADENQNVLEPKKQWEGAKVEPIGSGNSQPSAAYRINQNTIATKEEPTD--- 539
 QY 575 YAAIGKRVMEGVETEISGAVTPKQIHAGYSYLHSQIKTASNSRDGIFLLMKHSANLW 634

Db 540 YRSIGETIESKVELEIAISHLSDSVRLOAAATYTDIRKSSPOEGKRAVAYAPRQASAW 599
 QY 635 TTYQVTPRE-----LTIGGAVNAMSITSS-AGMHA-GGATFEDMAAYRFP-----KXKQ 684
 Db 600 LSTYVKSGLLEMLITLDSGIYVNGVTSDRLNTLTLSYTLVDVWVGVDLSIGLNSAQ 659
 QY 685 INADNIFNRHYARVGANTNIGSEKRTWTANLRYSF 722
 Db 660 LNVNINLIDKRYAAACNSL-SYCFGAERSIVGSVNAF 696

RESULT 37

AC0209
 ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
 C:Species: Nostoc sp.
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
 C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
 C:Accession: AC2079
 R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kultz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
 A:Reference number: AB1807; M01D:21595285; PMID:11759840
 A:Accession: AC2079
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-853 <KUR>
 A:Cross-references: GB:BA000019; PIDN:BAB73884.1; PID:g17131276; GSPDB:GN00179
 A:Experimental source: strain PCC 7120
 C:Genetics: A:Gene: alr2185
 C:Superfamily: ferrichrome-iron receptor 1; conB-dependent receptor amino-terminal homol

Query Match 11.1%; Score 420.5; DB 2; Length 853;
 Best Local Similarity 23.1%; Pred. No. 3.1e-21;
 Matches 187; Conservative 117; Mismatches 327; Indels 177; Gaps 30;

QY 3 QPMSVRINMTATVL--AALSSS---VF--AAQTADL-----ETVHIKGRSYN 45
 Db 135 QDANTTRVSLVGETILPTRELFPDSDEGLIFGLTSVTSALPPOQPEQSESPSAEPIE 194
 QY 46 AIVT-EKNGDYSSFAVTGKTPASLREIPQSVSITTNQVKDRNVDTQOLARKTPGLR 104
 Db 195 LVVTGKDGQVQNTATV--TRMTPIIDLPQSIQVPRQVLEQQTTRVDDALRNPGLV 253
 QY 105 VLSNDDGRSSVARGYVESEYNI--DGLPAQWQINGTLNLFAPDRVEMVMSGSGSLPDS 162
 Db 254 GSTNAETGNQITTRGFSTSNLPILRDGFRIYE--NFSQETSINLRIEVLKGPASVQYG 310
 QY 163 SGEMGIYVLRVRPRPKAQGHAAAFGTHKOYKAADVSGLSNDSYRGVMAQTVGA 222
 Db 311 QLDPGVINYLVTKPLSEPEYETIQAFQSGYGLIRPSFDVSGPLTDGKLYRLNA--TYOR 369
 QY 223 SPRAEKNRHEHETFAAAD--WDINPDTVL--GAGLYQORHLAPYNGLPADANNKLPSTL 278
 Db 370 EBFGRFPNTERETFFAPLSITMKISRTNVDSELELDSTR---PDTSLVAFGRVADY 426
 QY 279 P-OHVFVGAQDMNKFKMNSHDVPAADLKHVFNNGYGVKVMYSPRDADSNAPAGS----- 332
 Db 427 PYSRVENDPD-DEIDTKFSIAVNLHRPSSDWTLTANSFRYLOQDLFTQATLAGSINETT 485
 QY 333 KLGMRTPAGRPGCNTADDAACAVGLGTEIKOKALADASYSRPRFNGTANFVGADYN 392
 Db 486 GIITRTYAR-----EYKSDYSLSQTNVVGKFTTGSIKHTLLAAGVDN 528
 QY 393 R-----FRST-----NEGRTTYARGLALNEF 416
 Db 529 RGLDLDLVRRGRFTTININPVYGVPRRTDFTLPRATPFKKTETRLGTYLQDQILANN-- 587
 QY 417 RSTPQVDLIANARKGVARGVSHTVATENLDFGIYKGSFHPADGSLIGGRLGHYKIES 476
 Db 588 ---QFTVLAGLRVDTVDDEKDTFTDESKYD-----SAWSPVIGL----- 622


```
Qy 477 GEGKTLHKASKTKFTGYAGAVVLDNNDNSLYLSLSQLYTPQTNLDADGKLLKPRQGNQPE 536
Db 623 -----VKPVENLSIYTSRSFVPSFRDANGDFLOPERGAGYE 662
Qy 537 VGYKGYMDDRLNARVYFMRMKNKAAAPLNPNKKTRYAALGKRVMEGVETEISGAVTP 596
Db 663 VGKAEELGQNLFPATUAYFDITKQNVST-ADPDVLGAS-VATGEQSRSGIELSANGVTIAP 720
Qy 597 KWOIHAGYSYLHSQIKTASNRDDGIFL-----LMPKHSANLWTTYYQV-----TPELTIGG 647
Db 721 GWNIIAGYAYTDAEI-----SKONTIAGVNRLLPGTPKHSANLWTTVEIQKSLQGLGFL 775
Qy 648 GVNAMSGITSSAGMHAGGATFADMAAAYFT-----PKLKLQINADNIENRHHYARV 699
Db 776 GNYVY-----GKRFNGQNDFEVDSYFLTNAALFYRQNNRVGLNFNNIFDINY---I 825
Qy 700 GGANTFN-----IPGSERTWTANLRYSF 722
Db 826 SSAATLTRSTRSTEPGQPFVTVGSISLEF 853

RESULT 38
S74457
ferrichrome-iron receptor 3 - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1490
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Accession: S74457 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S74457
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-853 <KAN>
A:Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PTDN:BAAL6609.1; PID:g165168
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: fhuA_3
A:Start codon: GTG
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol
F:241-371/Domain: tonB-dependent receptor amino-terminal homology <TNN>
F:579-853/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 11.0%; Score 416.5; DB 2; Length 853;
Best Local Similarity 23.2%; Pred. No. 5.9e-21;
Matches 178; Conservative 116; Mismatches 311; Indels 161; Gaps 27;

Qy 10 INMTATVLAALSSVFAQTADLETVHKQGRSNAIYV-RKNGDYSFATVGTGTPKA 68
Db 165 VNATQSLVGLSPGKGVAADEED-----GNDAIQVVVVGQDEGAYVDDATTATLTD 217
Qy 69 SUREIPQSVSIITNQVKDRNVDTFDOLARKTPGLRVLNSDGRSSVYARGVEYSEYDID 128
Db 218 PLRDIPOSIVVQVQVLEDRQIRASEALQNVSGVQRGVGTGTSSEI-----FNIR 268
Qy 129 GLPAQMQSINGTLPNLFAF-----DRVEYMRGSPGLFSSGEMGGINVLVRKR 176
Db 269 G----FQPGGTLRDGFKFRDNFSIPTANLQRIEVLKGPASVLYGNLDPGGVINVTQK 324
Qy 177 PYKAFQGHAAAGETHKQYKAADVSGSLNSGSRGRV-MAQTGASPRPAKNNRHET 235
Db 325 PLSEPFYEVAMQAGNGLVRPTIDLSGLNSQRTALYRLNAAIEYGGNFRDPTDEVARFF 384
Qy 236 FYAAADWDINDPTVLGA--GYLYQQRHLAPYN-GLPADANNKLPSLPQHVFVGADWNKFK 292
Db 385 ISPVVTWQLSDQDRLRFEWDYLYDRR---PFRGIVA----- 418
```

```
Qy 293 MNSHDVFADLKHYPGNG-----GYCKVGMRYSDRDADSNYAPAGSKLGMKTPAGRPGCNTA 348
Db 419 -----FGTGADIPFDFRVLGELDFDARTNFS-AGYRLRHR-----FS 455
Qy 349 DDKACAVGLGTEIKQKALAPDASYSRPERL-----GNTANER-----VIGA 389
Db 456 DNWK-----LNRNFRPSYLDQAAEQTELVRDETGNLSRQFSRNEQQIRNVELQTDLLIG- 510
Qy 390 DYNRFRSTNEOGRITLYA-----RGLLALNEFRSIPQVDLI-----ANARKGV 432
Db 511 ---KF-STGPTQHTLLFGVDLSWQSAPPIFRGGVA-----APTINIFNPYGVIVAREPSI 560
Qy 433 RYSHTVATE-NLDFPGIYKASTFHPADGLSLGGRLGHYKIESGEGKTLHKASKTKFT 491
Db 561 NDFPDEVSESQTTGLFDQDQVTLTDNLKLMGGRFDTIDQSSSSNGESDERYDQAFS 620
Qy 492 GYAGAVDLDNNDNSLYLSQLYTPQTNLDADGKLLKPRQGNQPEVGYKGYMDDRLNAR 551
Db 621 PRLGLIVQPIEPVSLYASFSSRFQNFGRFDGSLLEVPFGTQYEVGVGRGEFLDGRLIAN 680
Qy 552 VSFYRMKDKNAAA--PLNPNKKTRYAALGKRVMEGVETEISGAVTPKWQIHAGYSYLHS 609
Db 681 LAAYEITVSNLAVTDPENPFS----IPSGEQRSGVEFDIAGEILPCWNIILASYATDA 736
Qy 610 QIKTASNRDDGIFLLMPKHSANLWTTYYQVTPELTIGGVNAMSGITSSAGMHAGGVATF 669
Db 737 RVTKDDNLEPCNLLEGVPFNSASLWSTYEQ-----AGDQLGLGFLGFLFVYGEROG 788
Qy 670 DAMAAYRFTPKLKLQINADNIFNR--HYARVGGGANTFNIPGSERT 713
Db 789 DLNNSFQIPSLYRTDI---SVFYRRNNWRAAINVNNLFNIDYIEAT 831

RESULT 39
AC2132
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2132
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.;
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2132
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-857 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74309.1; PID:g17131703; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2610
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 11.0%; Score 415.5; DB 2; Length 857;
Best Local Similarity 23.6%; Pred. No. 7e-21;
Matches 174; Conservative 106; Mismatches 296; Indels 161; Gaps 28;

Qy 47 IVTERKNGDYSFATVGTGTPKASIREIPQSVSIITNQVKDRNVDTFDOLARKTPGLRVL 106
Db 204 VVTGQODRYSVPNATATTATDTPLRDIPOSIAVPRQVLEDROQVIRASDALRSVSGVQOG 263
Qy 107 SNDDGRSSVYARGVEYSEYDIDGLPAQMQSINGTLPNLF-----ADPVEVMR 154
Db 264 NKVCGTSEV-----FNIRGFP-----QFGNLRGDFNNRNNFSIVETANLERIEVLK 310
Qy 155 GPSGLFPSSGEMGGINVLVRKRPTKAFQGHAAAGF--GTHKQYKAEADVSGSLNSGDSVR 212
Db 311 GPASVLYGNLDPGGVINFTVKQPLS--EPFYAAGLQVGSFGLVRPTLDLSGLPLNPERILL 368
Qy 213 GRVMAQTVGASPRPAEKNN---RHETFYAA--ADWDINDPTVLGAGLYQQRHLAPYN-G 266
```



```
Db 369 YRLNA-----AYEBCGNFRNFDTEVERFFISPVVTWKIGDRTDLRLLEY-SNDKRPYDRG 423
Qy 267 LPADANNKLPSPQHVFAVDWNNKFKANSHDVFAD-----LKHYFGNGGYKVKMR 317
Db 424 LVAFGNG-IADIPEDVLG-----BEDDSEKRTNPLAGRLHRFRPDMKLANQR 473
Qy 318 YSDRDAASVAPAGSKLGMKTPAGPCNTADDKACAVGLGTEIKOKALAFDASYSRPR 377
Db 474 YSSSDQTN-----RLBGRRLNETTGE-----LBRREFS 501
Qy 378 LGN-TNEEVIGADYNNRFRSTNEQRTTLA-----RGGLANFRSTIPVDLIANAR 429
Db 502 LBEFTFRNELQTDLVGNFATGSIQHTLLFGVDLSWVNDGGVSL--FEAPASIN-IFNPV 558
Qy 430 KGVRGYSHVATEENLDEF-----GIYKSTFHPADGLSLIGGRGKHIES 476
Db 559 YGI-----ATRPBRDFADVFPFGSQTDSIGVFVONQITLAEMLKLLVGGRRDNDQSS 612
Qy 477 GEGKTLHKASKTKFTGYAGAVYDLDNNNSLYLSQLYTPQTNLDADGKLLKPRQNGPE 536
Db 613 ASDERDQA-----FSPRVGIIVQPIEPISLYTSFSRSPQNFGRADGSLLEFVRGTYE 668
Qy 537 VGVKGSYMDRLNARVSFRMKDKMAA--PLNPNKKTIRYALGKRVNVEGVTEISGAV 594
Db 669 VGVRGFEFLNGSLITNLAAVEITRSNLAVTDPDPNPS---IPSGQRSGRGVGLDVTGQI 724
Qy 595 TPWKQIHAGYSYLHSQIKTASNSRDDGIFILMPKHSANLMTTYQVPELITGGVNAWSG 654
Db 725 LFGMNLIAAYATTDARVTDADDNLQPGNLADGVFPNSASLSTYEIO-----TGDPQ 776
Qy 655 ITSSAGMHAG-----YATPDMAAYRFTPKLKLQINADNIPNRHYA--- 697
Db 777 LFGGLGLFVYVGEHQGLNNSFQLPSYVRTDASIFYR-RNNWRAGININMLFNVDIEDSG 835
Qy 698 ----RVGANTRIPGS 710
Db 836 QRRNRINPGEFPTVRGT 852

RESULT 40
C95382
Probable ferrichrome-iron receptor [imported] - Sinorhizobium meliloti (strain 1021) mag
C|Species: Sinorhizobium meliloti
C|Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C|Accession: C95382
R|Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barclay-Hubler, F.; Bows,
, Kaiman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A|Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti
A|Reference number: A95262; MUID:21396509; PMID:11481432
A|Accession: C95382
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-714 <KUR>
A|Cross-references: GB:AE006469; PIDN:AAK65621.1; PID:914524105; GSPDB:GN00165
A|Experimental source: Strain 1021, megaplasmid pSymA
R|Galibert, F.; Flman, T.M.; Long, S.R.; Phihler, A.; Abola, P.; Ampe, F.; Barclay-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A|Authors: Kahn, D.; Kahn, M.L.; Kaiman, S.; Keating, D.H.; Kles, E.; Komp, C.; Lelaure,
hebaull, P.; Vanderbol, M.; Vorholter, F.U.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A|Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A|Reference number: A86039; MUID:21368234; PMID:11474104
A|Contents: annotation
A|Gene: SMA1747
A|Genome: plasmid

Query Match 10.9%; Score 411; DB 2; Length 714;
Best Local Similarity 24.6%; Pred. No. 1.1e-20;
Matches 187; Conservative 110; Mismatches 344; Indels 120; Gaps 30;
```

```
Qy 12 MTAATVLAALSSSVFAAQ-----TADLETYHIKQORSYNAIVTEKNQDYSSFAVTVGT 64
Db 24 LIGCTAFALTLALSPAOAVPEGDTTVLETIVAHGAGGSVLTNDEDSKSIATRTGA 83
Qy 65 -KIPASIREIPQSVSITNQVKRNVDPFDQARTPL--RVLSNDGRSSVARGYE 121
Db 84 GKMPFDILAPASVSVITSEKIEERAADTIEQVQVTAQVTVDPFYSDDRYDFYDRGPT 143
Qy 122 YSEFNIDGLPAQMOSINGLPLNLPAPDRVEMRGPGGLFDSGEMGIYNLVKRPRTKAF 181
Db 144 PTTYR-DGL-AIGRTPAVGRPEPTAFERILEVKGASSSFGAAEPGGSVNVYTKTPKSR 201
Qy 182 OGHAAGFPTHKQYKAEADVSGSLNSDSGVRGVAQTGASPRPAEKNNRHETFYAAD 241
Db 202 FGEVGTGGSFSHKELGDFGDNLTADETL-----SYRLGKQQRSDAEVDYSG 250
Qy 242 WDN-----PDTVLGAGTLYQQRHLAPYNGLPADANNKL-PSLPQHVFGAD--W 288
Db 251 DDENFVMGVTWRPTDTSITFEIFD--HL-DKQVPGSGHPLGTDPFDRDQFGEEDYYF 307
Qy 289 NKFRKNSHDVPAADLKHYPNGGKVKVMYSDRDAASNYAFAGSKLGMKTPAGRPCNTA 348
Db 308 SETNRSISVLPD--HDFGNGLSFSNARYSNUNGFGSAITGS-----TP 351
Qy 349 DPKACAVG--LGTEIKOKALAFDASYSRPFRLGNTANEVIGADYNNRFRSTNEQRTTL 405
Db 352 TGGSYVAGRYFPFNGEKSTQPIDALVYEAIDLNVESITLFGADNKKESD----- 403
Qy 406 YARGGLANFRSIPVDLIANARKGVRSYSHVATENIDE--FGIYKSTFHPADGLSL 463
Db 404 -----SANFYAPAPSIDMEDPIYSGGPGAMA PYASTNNDQOTNAIYLQODLTFEFDKTV 457
Qy 464 IGGG-----LGHYKESGEKTLHKASKTKFTGYAGAVYDLDNNNSLYLSQLYTPQT 518
Db 458 SFGLRNDWLDLSETNLAA--TRPAGNRBEPTTRIGASYKTEELAPYISAESAP-- 512
Qy 519 NLDADGKLLKPRQNGQFEGYKGSYMDRLNA--RVSFRMKDKN-----AAAPLNPNK 571
Db 513 --PAAGS--DPTTGQYEVGIR--YRPDAFPAMFTASVVDLTKGNTVDPQVYTLPGTYE 566
Qy 572 KTRVYALGKRVNVEGVTEISGAVTPKQIHAGYSYLHSQIKTASNSRDDGIFILMPKSHA 631
Db 567 KYRH-----RGFELKAKAVTNNISVIAAYSITDSKIEBPGCANDGNRLMRVPKMA 618
Qy 632 NLMTYQVPELITGS--GVA-----MSGITSSAGNHAGYATFDMMAAYRFTPKK 682
Db 619 SWGTYTTLLEGDARGGMLFGLGARVTDAYTISITNTSSEA--VVFDAAFYTKIQENTT 676
Qy 683 LGINADNIFN-RHYVAVGANTFNPISGERTMTANLRYSF 722
Db 677 FQNLNNLFDKXHVASKDSGAVYVN-PG--RSILATLRQSW 714

RESULT 41
C81861
Hypothetical protein NMA1663 [imported] - Neisseria meningitidis (strain Z2491 serogroup
C|Species: Neisseria meningitidis
C|Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C|Accession: C81861
R|Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
, Holroyd, S.; Jagsis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A|Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
A|Reference number: A81775; MUID:20222556; PMID:10761919
A|Accession: C81861
A|Status: preliminary
A|Molecule type: DNA
A|Residues: 1-635 <PAR>
A|Cross-references: GB:AL162756; GB:AL157959; NID:97380091; PIDN:CAB84891.1; PID:9738030
A|Experimental source: serogroup A, strain Z2491
A|Gene: NMA1663
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 18, 2002, 06:37:39 ; Search time 14.3071 Seconds
(without alignments)
2093.073 Million cell updates/sec

Title: us-09-889-267-2

Perfect score: 3776
Sequence: 1 MGQENSVPRIMTATVLA.....NTFNIGSERTWTANLRYSP 722

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	828.5	21.9	826	1	PBUA_PSESP
2	821.5	21.8	815	1	PFVA_PSEAE
3	767.5	20.3	819	1	PUPA_PSEPU
4	737.5	18.5	729	1	FHUE_ECOLI
5	715.5	18.9	720	1	PPTA_PSEAE
6	708.5	18.8	809	1	PUPB_PSEPU
7	470	12.4	760	1	YBIL_ECOLI
8	423	11.2	702	1	FOXA_YEREN
9	401.5	10.6	710	1	FOXA_YEREN
10	387.5	10.3	747	1	FHUA_ECOLI
11	374	9.9	735	1	FCU_YEREN
12	303	8.0	673	1	FCU_YEREN
13	300	7.9	673	1	FCU_YEREN
14	267	7.1	663	1	CIRA_ECOLI
15	256.5	6.8	746	1	PEEA_PSEAE
16	221.5	5.9	345	1	YEEA_HAEIN
17	215.5	5.7	726	1	FATA_VIBAN
18	214	5.7	652	1	IRGA_VIBCH
19	214	5.7	700	1	YNCB_ECOLI
20	205	5.4	614	1	BTUB_SALTY
21	205	5.4	774	1	FECA_ECOLI
22	200	5.3	687	1	HEMR_YEREN
23	198	5.2	732	1	IUTA_ECOLI
24	195	5.2	676	1	HMUR_YERPE
25	193.5	5.1	746	1	FEPA_ECOLI
26	193.5	5.1	746	1	RHTA_RHIME
27	191	5.1	614	1	BTUB_ECOLI
28	190.5	5.0	758	1	FCUA_YEREN
29	172	4.6	915	1	TBPI_NEIGO
30	167.5	4.4	908	1	TB12_NEIMB
31	165.5	4.4	725	1	HXC2_HAEIN
32	162.5	4.3	810	1	HPUB_NEIMA
33	161	4.3	911	1	TB11_NEIMB

34	158.5	4.2	723	1	Y262_HAEIN
35	157.5	4.2	810	1	HPUB_NEIMC
36	143	3.8	2003	1	YDBA_ECOLI
37	140.5	3.7	744	1	HXC1_HAEIN
38	137.5	3.6	1061	1	OAR_MYXXA
39	137	3.6	687	1	VITA_VIBCH
40	137	3.6	2334	1	WAPA_BACSU
41	136	3.6	993	1	HGBC_HAEIN
42	130.5	3.5	4349	1	FAT2_HUMAN
43	129	3.4	699	1	CH11_BACCI
44	129	3.4	1046	1	HHUA_HAEIN
45	129	3.4	1462	1	GTFD_STRMU

ALIGNMENTS

RESULT 1	ID	PBUA_PSESP	STANDARD;	PRT;	826 AA.
AC	008017;				
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	Ferric-pseudobactin M14 precursor.				
GN	PBUA.				
OS	Pseudomonas sp. (strain M14).				
OC	Bacteria; Proteobacteria.				
OX	NCBI_TaxID=306;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=94104606; PubMed=8277948;				
RA	Morris J., Donnelly D.F., O'Neill E., McConnell F., O'Gara F.;				
RT	"Nucleotide sequence analysis and potential environmental				
RT	distribution of a ferric pseudobactin receptor gene of Pseudomonas				
RT	sp. strain M14."				
RL	Mol. Gen. Genet. 242:9-16(1994).				
CC	- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC				
CC	PSEUDOBACTIN M14.				
CC	- SUBCELLULAR LOCATION: Outer membrane.				
CC	- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.				
CC	- This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; X73412; CAA51812.1; -				
DR	InterPro; IPR000531; TonB_boxC.				
DR	PFam; PF00593; TonB_boxC; 1.				
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.				
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.				
KW	Outer membrane; Iron transport; Transport; TonB box; signal; Receptor.				
FT	SIGNAL	1	44		
FT	CHAIN	1	826		
FT	SITE	110	119		
FT	SITE	809	826		
FT	SEQUENCE	826 AA;	90388 MW;	5D440D6AC7974CB CRC64;	
Query Match	21.9%;	Score 828.5;	DB 1;	Length 826;	
Best Local Similarity	29.9%;	Pred. No. 2.8e-49;			
Matches 227;	Conservative 129;	Mismatches 303;	Indels 101;	Gaps 26;	
OY	16	TVALALSSVFAQADLETVHIIKGRSYNAIVTEKNGDYSSFAVTVGTRIPASIREIQ	75		
DB	115	TVLGPATGSAM-----ELAPTVNMSRL--GATTEGNSYTTGGVITIGGV-HSLKETPQ	166		
OY	76	SVSITNQGVKDRNDVTGPQALARKTPGLAVLSNDDGRSSVYARGVYS-ENYNDGLP---	131		
DB	167	SVTVVTRKMLDDQDNTITIQVMEKTPGIVYDPSPMGKXYFSGFRMSGOYQDGVLDI	226		


```

QY 132 -----AQMOSINGTLNLFAPDRVEMVRGSPGLFSSGEMGGIVNLRKRPYKAFQGHAAA 187
DB 227 GSSYVQADSFNS---DMAIYDRVEVLGAAGMMKAGAGTAGGVNFRKRGQDTAHTQLSL 283
QY 188 GFTGHQKQKAEADVSGSLNSDCSVGR--VMAQT-----VCGASPRPAEKNNRHTEFYAA 239
DB 284 SAGTWNVRGQVDTGGPLNDSGTIIGRAVVTETQYFYDVG-----SKDQIYYGA 335
QY 240 ADWDINPDTVLGAGLYQORHLAP--YNGLPADANNKLPSLPQHFVFGADNMKFKNNSHDV 298
DB 336 LDFDLSPDTTLGLGPAWEDVDATPCWGLPRYADGSDLHLKSTCLNTAWNQRKSRATY 395
QY 299 FADLKHFGNGYGVKVMRYSRDDSDNSVAFAGSKLGMKTPAGRGCCNTADKCAVGLG 358
DB 396 FADLKHQFNDDSLVAGVYSRNTQMEYAFSPGAV---PVGATATNT-----LMLGSI 446
QY 359 TEIKOKALAFDASRPPRLGNTANFEFVIGADYNFRSTNEQGRITLVARGGL-----AL 413
DB 447 YDQDQYDFDAYVDGKDFAPGQOQHELTIGANASR-----SHKDDFYAVALPQRQNVL 500
QY 414 NEFRSIPQVD---LIANARKGVGRGYSHTVATENLDEFGLYKSTFHPADGLSLIGGRLG 470
DB 501 DRNHHIPQDESYLANASRGCPVDH-----IKQYGAYSIALKLADPLTLVLGSRVS 554
QY 471 HYKIES-----GEGKTLH-KASKT-KFTGYAGAVYDLNNSLYLSQLYTPO--TNL 520
DB 555 WYKSDTDSVOYFRGEGTQVDTKSTGTQVTPPAGVLFDLNLTAYASVDTFTPGAYK 614
QY 521 DAGKLLPRQNGQEVGVKGYSDMDRLNARVSFVRMKDKNA-----APLNPNNK 572
DB 615 TIDGSTLPLVQGVSELGKGFGLNLSNLTNLTFTLQKDAQDDPCEDSSCSINSK 674
QY 573 TRYAALGRKVRMGVETEISGATPKWQIHAGYSYLHSQIKTASNSRDDGIFL--LMPKHS 630
DB 675 VR-----AGCFRAEVSSEVIDRLQLLAGYTYTQTKVLEADATQDGVVYNSVPRHL 726
QY 631 ANLWTYQVTPEL---TIGGVNAMS--ITSAG---MHAGYATFDMAAYRTPPKLK 682
DB 727 LRWGDYLSGLPLDRVTIGAGVNAQTGNVRTSPIGGDNIDGAGYAVWNGRIGRYIDTWS 786
QY 683 LOINADNIFNRHYARVGAGTANFNIPGSERTWTANLRYSF 722
DB 787 VALNGNPLDKRYYSTIGTGFNGFYGDPRNFVMSVKADP 826

RESULT 2
FPVA_PSEAE STANDARD; PRT; 815 AA.
AC 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferripyoverdine receptor precursor.
GN FPVA OR PA2396.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OX NCBI_TaxID=287;
RN [1]_TaxID=287;
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.
RC STRAIN=CD10;
RX MEDLINE=93328663; PubMed=8335619;
RA Poole K., Neshat S., Krebs K., Heinrichs D.E.;
RT "Cloning and nucleotide sequence analysis of the ferripyoverdine
receptor gene fpva of Pseudomonas aeruginosa.";
RL J. Bacteriol. 175:4597-4604 (1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Lamont I.L.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
[3]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=2043733; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
Hickey R.L., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -1- FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L10210; AAA25819.1; -.
CC EMBL; U07359; AAB60199.1; -.
CC EMBL; AE004666; AAG05786.1; -.
CC PIR; A40601; A40601.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC.1.
CC PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE_NEG.
CC PROSITE; PS01156; TONB DEPENDENT REC 2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 815 FERRIPYOVERDINE RECEPTOR.
FT SITE 798 815 TONB C-TERMINAL BOX.
FT CONFLICT 716 716 Y -> F (IN REF. 1).
FT CONFLICT 745 746 MISSING (IN REF. 1).
FT CONFLICT 750 750 S -> R (IN REF. 1).
SQ SEQUENCE 815 AA; 91167 MW; A11AE4AA5A290F35 CRC64;

Query Match 21.8%; Score 821.5; DB 1; Length 815;
Best Local Similarity 29.0%; Pred. No. 8.3e-49;
Matches 217; Conservative 145; Mismatches 312; Indels 75; Gaps 24;

QY 8 PRINMTATVLAALSSSVFAAQATADLETVHKQORSYNAIVTEKNGDYSSFAVTVGTKIP 67
DB 108 FQGNATITSVAEADSSV-----DLGATMTITSNQL--GTITEDSGSYTPGTIATRLV 159
QY 68 ASLREIPQSVSIITNQVKORNVDTFDQARLKTPLGLVLSNDGRSSVYARGVESEYNI 127
DB 160 LTPRETPOSITVTTRQNMDDFGLNIDVDMRHPTGITSAYDTRNNYARGFSLNNFYQ 219
QY 128 DGLPAQMQSI---NGTLPLNLFAPDRVEMVRGSPGLFSSGEMGGIVNLRKRPYKAFQ 183
DB 220 DGIPTARNVGYSGAGNTLSDMAIYDRVEVLKAGATGLLTGAGSLGATINLRKPTHEFK 279
QY 184 HAAAGFGTHQYKABADVSGSLNSDCSVGRVMAQTGVGASPRPAEKNNRHTEFYAAADWD 243
DB 280 HVELGAGSDWNRSELDVSGPLTESGNVRGRAVAAYQDKHSFMDHYERKTSVYTGILEFD 339
QY 244 INPDTVLGAGLYQORHL--APYNG-LPA--DANNKLPSLPQHFVFGADNMKFKNNSHDV 299
DB 340 LNPDTMLTVGADYQNDPKSGWGSFPLFDSQGNRNDVSRFSNNGAKWSSWEQVTRTVF 399
QY 300 ADLKHFGNGYGVKVMRYSRDDSDNSVAFAGSKLGMKTPAGRGCCNTADKCAVGLGT 359
DB 400 ANLEHNFWANGVGVQL--DHKINGYHAPLGAIMG-DWPA-----PDNSAKIVAQKY 448

```



```

OY 360 EIKOALAFDASYSRPFPLGTANAEFYIG--ADVNRFRSTBEOERTLVARG-GLALNEF 416
Db 449 TGETKSNLSLDLYLTGPFLCERHELVAIGSASHW----EKGSTWMLNRNDNTTDDF 503
OY 417 RSIPOVDLIANARKVGVSHTVA TENLDERGIGKSTFHPADLSLIGGRLCGHYIES 476
Db 504 -----IMNDGIGKRPMDTPEQYIDDKTRQUGSFMTRAFNVTDDLNLFLGGRVYDRV-T 557
OY 477 GEGKTLHKASTKFTGTAGAVAYDINDNNSLYLSLSQLYTPQTN--LDADGKLLKPRQNG 534
Db 558 GLNPITRESG--RFPIYGAAYDINDTYSVASYATDIFPMQDSWYRDSNKLLEPDEQON 615
OY 535 FEVGVKSSYMDRLNARVSFRMDKXAAAP-----LNPNNKTRRYAALG-KRVMBEVERT 568
Db 616 YEIGKKGYYLDGRINTSLAYEIHENKRAEDBALYNKSPFNPAITYAYKGIKAKTKYEYA 675
OY 589 EISGAVTEKMOIHAGYSYLHSQIKTASNSRBDGIFLMPKHSANLMTTYO---VTPELTI 645
Db 676 EISGELIAGWQVQAGYT--HKIINDDSGKK--VSTHEPDQJLSLYTSYFPGALDKLIV 730
OY 646 GGGV-----NMSGITSAGMHAAGVATFDMAAYRFTPLKLIQINADNIENR 693
Db 731 GGGARWQKSSQWVYNNPRSRWEK---FSQGDYLVLDLMARYQTIDTSLASVANNVFDK 787
OY 694 HYIARVGCANTFNIPGSERTTANLARSF 722
Db 788 TYTNIGFYTSASY-GDPRNLMFSTRMDF 815

```

RESULT 3			
PUPA_PSEPU	STANDARD:	PRT:	819 AA.
ID	PUPA_PSEPU		
AC	P25184;		
DT	01-MAY-1992 (Rel. 22, Created)		
DT	01-MAY-1992 (Rel. 22, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	Ferric-pseudobactin 358 receptor precursor.		
GN	PUPA.		
OS	Pseudomonas putida.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;		
OC	Pseudomonas.		
OX	NCBI_TaxID=303;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MCS358;		
RX	MEDLINE=91260449; PubMed=1646376;		
RA	Bitter W., Marruga J.D., de Weger L.A., Tommassen J., Weisbeek P.J.;		
RT	"The ferric pseudobactin receptor Pupa of Pseudomonas putida MCS358:		
RT	homology to TonB-dependent Escherichia coli receptors and specificity		
RL	of the protein.";		
RL	Mol. Microbiol. 5:647-655 (1991).		
CC	-1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC		
CC	PSEUDOBACTIN 358.		
CC	-1- SUBCELLULAR LOCATION: Outer membrane.		
CC	-1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; X56605; CAA39942.1; ..		
DR	PIR; S15169; S15169.		
DR	InterPro; IPR000531; TonB_boxc.		
DR	Pfam; PF00593; TonB_boxc; 1.		
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.		
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.		
KM	Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.		
FT	SIGNAL 1 47		
FT	POTENTIAL.		
FT	FERRIC-PSEUDOBACTIN 358 RECEPTOR.		

FT	115	122	TONB BOX.
FT	802	819	TONB C-TERMINAL BOX.
SO	SEQUENCE	819 AA; 91015 MM; 1B36164F86207951	CNC64;
Query Match		20.3%;	Score 767.5; DB: 1; Length 819;
Best Local Similarity		27.8%;	Pred. NO. 4.36-45;
Matches 214; Conservative		134; Mismatches 306;	Indels 117; Gaps 26

Qy	8	FRMTAAATVTLAALSSVFPAAGQADLAEITHIKQREYNALVTEKNDGSSFAVYVGTGKIP	67
Db	110	YQIDGNTVTVTASL-----AAKQGIELSATNVNSIGLSETTGTSYTRVTRATKKN	164
Qy	68	ASLSEIPQSVSIITNQCKVRNDVTFDOLARKTPGLRVLSSNDG--RSSVYARGYSEYXN	126
Db	165	LSIEFTQTITVTVTRQMDQDHGSSNNEVLTQPPGI--TMSQGGKEFNIYNSGSAINIQQ	223
Qy	127	IDGLPA-----QMSINSTLPNLFAPRVEVMRGSGSLPSSGEMGIVNLVLRPKPAQO	182
Db	224	FDGVTTVQDNQOTNRMSPTLMDVGLAYRIEIVRGATGLMTAGGPPSAVVNVIRKRPREFX	283
Qy	183	GHAAGGCHKQYKAAEDVGSLSNDGSSVRGRMAQTGASPRPAKNNR-----H	233
Db	284	SHIQAGSGSDVYRAEADVSGPLTDCGRVRGRFA-----AAQDNHTFMDWYDOR	334
Qy	234	ETFPAAADWPIINDPTVLGAGYLQOEBHLAPYNLPA-----DANNKLPSPLOHFVAGAWN	289
Db	335	DVLXGVVEADVDTOTVARFGI---DGQTKVNGAPQVPIIYTNQGPTNRSRSTSSPAR-	390
Qy	290	KEKKNSHDVFADLKHVFG-----NGGYGKVRYSDDRDASNYPAGSKLGMKTPAGRP	343
Db	391	-----GYDDPTTNTYTFEGLEQOLAHMOWEFLAAAYMDVDRDSSSY-----YSTTNS	439
Qy	344	GCMTADDKAGAVGIGETIKQKALAFDASYSRPRCLNTANEPIGADVNRFRSTNQRT	403
Db	440	YLELDGSTIEISAGIVT-AKHQKQVDATLQGPQLGQTHELIVGYNYLEYENKH----	493
Qy	404	TLVARGGLAINEFRSIPQVDL-----IANARKG-----VRGSHYVATENLDEFGYKRS	453
Db	494	-----RGD-----SGPDVAINFYMDNDQTPKRGDDEIIRGIQVINSNR-----QSGVFVAS	539
Qy	454	TEHPADGLSLIGGRIGHYKIE-----SGEGKTLHKASKTKPTGYAGAVIDLDNNSLY	507
Db	540	RFNLTDLHLILGARASNYRFYALWIRIGNEPAPYKQVBERGVTPAAGIVYDLTNEQSY	599
Qy	508	LSLSQLTPTQTNLDAGOKLLKPRQNGQFVGYKGSIMDRDLNARSFYMKKKNAAAPLN	567
Db	600	ASYVDIEPKPNNAVDTGKPLDPEVGNKYEIGWGEFLLEGRLNANIALYVWKKDNLAESTN	659
Qy	568	---PNNKRTKTYAALGKRVMEGET-----EISGAVTPKKQIHAGYSYLHSQIKTASNSHD	619
Db	660	EVPDSC-----GLIASANDGATTKGVDELSEVLPGNNVFTGIS--HTRIEDADGR-	712
Qy	620	DGIFLMPKHSANLWTTYQVTPF--LTTGGGVNANSGITSSAGMH-----AGYATPDA	671
Db	713	--LTPQLPMDTFPFMWYTYRLPGEMEKTLTGCGGVNMSKSTLNFARNSHVTODDVFVTSL	770
Qy	672	MAARFPFKLQIANDNIENRHYVARVCGANFNIPGSERTWTANLRSPF	722
Db	771	MARYRINESLAATLVNINIYFDKRYA--GWAAGSYGHYGAPRNATVYLKYDP	819

RESULT 4

PHUE_ECOLI			
ID	PHUE_ECOLI	STANDARD;	PRT; 729 AA.
AC	P16869;	P77292;	
DT	01-AUG-1990 (Rel. 15, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	PhuE receptor precursor (Outer-membrane receptor for Fe(III)-coprogen		
DE	Fe(III)-ferritorxamine B and Fe(III)-rhodotulic acid).		
GN	PHUE OR B102.		
GN	Escherichia coli.		
DC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;		

OC Escherichia.
OX NCBI_TaxID=562;
(1)
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90286919; PubMed=2162465;
RA Sauer U., Hantke K., Braun V.;
RT "Sequence of the fhuB outer-membrane receptor gene of Escherichia
col K12 and properties of mutants."; Mol. Microbiol. 4:427-437(1990).
(2)
RN SEQUENCE FROM N.A.
RP STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Sampai G., Seki Y., Horiuchi T.;
RA "The complete genome sequence of Escherichia coli K-12."; Science 277:1453-1474(1997).
(3)
RN SEQUENCE FROM N.A.
RP STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
Sampai G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
Yano M., Horiuchi T.;
RA "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map."; DNA Res. 3:137-155(1996).
(4)
RN SEQUENCE OF 37-50.
RC STRAIN=K12;
RX MEDLINE=87194585; PubMed=3032906;
RA Sauer M., Hantke K., Braun V.;
RT "Ferric-coprogen receptor FhuE of Escherichia coli: processing and
sequence common to all TonB-dependent outer membrane receptor
proteins."; J. Bacteriol. 169:2044-2049(1987).
CC -!- FUNCTION: REQUIRED FOR THE UPTAKE OF IRON(III) VIA COPROGEN,
FERRIOXAMINE B, AND RHODOTORULIC ACID.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
BE ACTIVE.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; X17615; CAA35616.1; -;
DR EMBL; AE000210; AAC74186.1; -;
DR EMBL; D90745; BAA35809.1; -;
DR EMBL; D90746; BAA35917.1; -;
DR PIR; S09262; S09262.
DR PIR; A26875; A26875.
DR EcoGene; EG10306; fhuE.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transceptor; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 36
FT CHAIN 37 729 FHUE RECEPTOR.
FT SITE 42 49 TONB BOX.

FT SITE 712 729 TONB C-TERMINAL BOX.
FT MUTAGEN 44 44 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 46 46 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 163 163 G->A,V,N,D,H: IMPAIRED TRANSPORT AND
GROWTH.
FT CONFLICT 363 363 G->C (IN REF. 1).
SQ SEQUENCE 729 AA; 81232 MW; 68ACEE7D110F76CD CRC64;
Query Match 19.5%; Score 737.5; DB 1; Length 729;
Best Local Similarity 27.3%; Pred. No. 4.2e-43;
Matches 211; Conservative 148; Mismatches 308; Indels 105; Gaps 26;
Qy 3 QPMSVFRINMTAAV-IAALSSSVFAAQAQADLETVHVKQBSYNAIVTEKNGDVSSEFAVT 61
Db 11 QYQAITPSSLGAGIALALLPSAFAAPATE-ETVIVEGSAT---APDDGNDYSVTSTS 66
Qy 62 VGTIPASLRIPQSVSIITNQVKDRNVDTFDOLARKTGLRLVSLNDDGRSSVYARVE 121
Db 67 AGTKMQWTDIPQSVTIVSQORMEDQQLQTLGEVMENTLGISKQSDSDRALYISRGFQ 126
Qy 122 YSEYNIDGLPAQMS---INGTLPLNLFADRVEMVRGPGSLFSDSGEMGGIVNLVRKPT 178
Db 127 IDNYMVDGIPYFESRWNLGDALSDMALFERVEVVGATGLMTGTGNPSAAINVRKHAT 186
Qy 179 -KAFQGHAAAGFGTHKQYKABADVSGSLNSDGSVRGRVMAQTGCA---SPRAEKNRHE 234
Db 187 SREFKGDVSAEYSGWNERVYVADLQSLPTEDGKIRARI---VGGYQNDSDWLDRYNSEK 242
Qy 235 TFYAA-ADWDINPDTVLGAGLYQOORHL--APYNGLPA-DANNKLPSPQHVFGADWNK 290
Db 243 TFFSGIVDADLGDUTLTSAGYEYQRIQDVNSTWGLFRWNTDSSNSVDRASTAPDWAY 302
Qy 291 FKMNSHDVFDLKHVFGNGGYGKVMRYSDRDADSNY----AFAGSKLGMKTP----- 339
Db 303 NDKEINKVEMTLKQFPADTQATLNATHSEVEFDSKMYVDAYVYKADGMLVGVSNGP 362
Qy 340 ----AGRPCCNTADDKACAVGLGTE-----IKKALAFDASYSRPFR--LGNTANEF-- 385
Db 363 GFDYVGGTGWNSGKRKYDADLDFADGSYELFGRQHNLMFGSGYSKQNNRYFSSWNIIPD 422
Qy 386 VIGADYNRFRSTNEQGRITLLYARGGLALNEPRSIPOVDLIANARKVGRGYSHTVATENLD 445
Db 423 EIGSFYN-FNG-----NFPQTD-----WSPQSLAQDDTT 450
Qy 446 EF-GIYKSTPHPADGLSLIGGRLGHYKESGEGKTLHKASKTKFTGYAGAVYDLNDNN 504
Db 451 HMKSLYAATRVTLADPLHLILGARVYTNRVDT---LTYSMKKNHTTPYAGLVFDINDNW 506
Qy 505 SLYLSLSQLYTPQTNLDADGKLLKPRQNGQEVGYKSGYMDRLNARVSFYRMKDKNA- 563
Db 507 STYASYTSIFQPNDRDSSGKYLAPITGNNYELGLKSDMNSRLTTTLAIFRIEQDNVAQ 566
Qy 564 ---APLNPNNKKTVAALGKRVMGVEVETELSGVTPKQIHAGYS-YLHSQIKTKASNRD 619
Db 567 STGPIPGSNGETAYKAVDGTVSGVFEELNGALTDWQLTFGATRYI-----AEDNEG 620
Qy 620 DGIFLLMPKHSANLWTTYQ--VTPELTIGGVNAMSGITSSA-----GMHAGGYATFD 670
Db 621 NAVNPNLPTTKMFTSYRLPVMPELTVGGVNNQNRVYTDVTPYGTFRAEQGSYALVD 680
Qy 671 AMAAYRTPKLLQINADNINRHHYARVGANTFNIPGSGERTWTANLRYSF 722
Db 681 LFTRYQVTKNFSLOGNVNLFDKTYDYNVEGSIYV---GTPRNFSGITGYQF 729
RESULT 5
FPTA_PSEAE
ID FPTA_PSEAE STANDARD; PRT; 720 AA.
AC P42512;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fe(III)-pyochelin receptor precursor.

GN PPTA OR PA4221.
 OC Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 39-51.
 RC STRAIN=PAO / 1A602;
 RX MEDLINE=94117363; PubMed=8288523;
 RA Ankenbauer R.G., Quan H.N.;
 RT "FpA, the Fe(III)-pyochelin receptor of Pseudomonas aeruginosa: a
 phenolate siderophore receptor homologous to hydroxamate siderophore
 receptors.";
 RL J. Bacteriol. 176:307-319 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAO1;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Huffnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gaudier R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
 opportunistic pathogen.";
 RT Nature 406:959-964 (2000).
 CC -1- FUNCTION: HIGH-AFFINITY OUTER MEMBRANE RECEPTOR REQUIRED FOR THE
 -1- TRANSPORT OF Fe(III)-PYOCHELIN.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DR EMBL, U03161; AAC3213.1; -
 DR EMBL, AE004839; AAG07609.1; -
 DR InterPro, IPR000531; TONB_boxc.
 DR Pfam, PF00593; TONB_boxc.1.
 DR PROSITE, PS00430; TONB-DEPENDENT_REC_1; FALSE_NEG.
 DR PROSITE, PS01156; TONB-DEPENDENT_REC_2; 1.
 KM Outer membrane; Iron transport; Transport; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 38
 FT CHAIN 39 720
 FT SITE 703 720 TONB C-TERMINAL BOX.
 SQ SEQUENCE 720 AA; 79992 MW; DA796313116B0CC2 CR664;
 Query Match 18.9%; Score 715.5; DB 1; Length 720;
 Best Local Similarity 27.2%; Pred. No. 1.3e-41;
 Matches 200; Conservative 137; Mismatches 323; Indels 75; Gaps 20;
 QY 21 LSSVRAAQTAD-----LFTVHIKGGRSYNAITENKNGDSSRAVYVGTGKIPSLSR 72
 DB 28 LALSLPAAVADARKDGETELPDMVSGS--STSAIQPG-----VTTLGAVPLKPR 78
 QY 73 IPQSVIITNOQVKNVDFTDOLARKTGLRLVLSNDGSSVYARGYSEBNIGLPA 132
 DB 79 LQGSASVIERHERLEQONTLSLDEAMQATGVYQFQLTLTAYVAGFKVDSLELQVRA 138
 QY 133 QMOSINGTLPNLFAPDREVEVRGSGSLPSSGEMGIVLVIRKRPKAFQGHAAAGFTH 192
 DB 139 LIGNTASSQDMVAIVERVELIRSGNGLHGTGNPATVNLVRKRPQREFAASTTLISGRM 198
 QY 193 KQYKARADVSGSLNSDGSVGRMAQTGASPRP-----AEKNRHETFFAADMINDP 247
 DB 199 DKYRAEDVVGGLSASGNVRGRAVA---AYEDRDYVDVADQSTR--LITGVTEFLSDP 253

QY 248 TVLGAGYLYQO-RHLAPVINGLPADANNKLPSPLOHVFVADNKKFRMNSHDVFAIDKHYF 306
 DB 254 TLITVGAQYQHDISTITNMAGVPMADGSLGSRDYLVDWDFRFXMDTYRAFGSLEQOL 313
 QY 307 GNGGYGVGMRSDDADSDNVAFAQSKLGMKTPAPRPGCNTADDAKACANGLETETIQKAL 366
 DB 314 GGGWKGKVAEY--DEADSLRYASFGAIDPOTG-----DGGQLWGAAYKFGSIOR 363
 QY 367 APDASYSRPRRLGNTANEFVIGADYRFRSTNEQGRITLYARGALANFRSIPQ--VDL 424
 DB 364 SLDANINGVRLPGLTHELIGGVTAQGETRODTAR-----FNLNPNPVAIV 410
 QY 425 IANARKVR---GSHVATENLDEFGYKSTPHADGLSLIGGRIGHYKIESBECK 480
 DB 411 YRWDPHGVPRPQIGQYTSPTGTTTQKGLYALGRITAEPLTLVVGVRESGW--DDDTPA 468
 QY 481 TLHKASKTFPGYAAVYLDNNSLYLSLSQLYTPQNTLDA-DGTLKLRPGNGPEVGY 539
 DB 469 TPRKPR-OFTPEYGLIWPARDWSVYSAVYOPQADROGTWNSPPLSPVGKTYETGI 527
 QY 540 KGSYMDRLNARVSFYRMKNAAA-----PLPNKKTRVALGKRVMEGVETEISGAV 594
 DB 528 KGEIADGRNLGLAARLIDLENNPQEDPHRPPNN--PFYISGKVRSGFELBGTGL 585
 QY 595 TPRMOIHAGSYLHSG-IKTAASNSRDDGIFLLMPKHSANLWTTYQV--TPELLTGGVN 650
 DB 586 TPYWSLSAGYTTSTFYLDSQNSDSTRYSTFTPHLLRLKLSNVYDLPWQDRWSVGGGLQ 645
 QY 651 ANSGIT---SSAGMHAAGYATPDMAAAYFTFKLQIADNIFNHHYARVAGANTFNI 707
 DB 646 AOSDYSVDYRGVSMRGQGYALVNMRLGYKIDHMTAAVAVNNVLFRTYQYSLSPNMNMR 705
 QY 708 PSERTTANLRYSF 722
 DB 706 YGEPNSFNVISLGAIF 720
 RESULT 6
 PUBS_PSEPU
 ID PUBS_PSEPU STANDARD; PRT; 809 AA.
 AC P38047;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Ferric-pseudobactin Bm7/Bn8 receptor precursor.
 GN PUPB.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WCS358;
 RX MEDLINE=93116856; PubMed=8392140;
 RA Koster M., van de Vosseberg J., Jeong J., Weisbeek P.J.;
 RT "Identification and characterization of the pupB gene encoding an
 RT inducible ferric-pseudobactin receptor of Pseudomonas putida
 WCS358.";
 RL Mol. Microbiol. 8:591-601 (1993).
 CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORES FERRIC
 CC PSEUDOBACTIN Bn8 AND FERRIC PSEUDOBACTIN Bm7, IRON CHELATING
 CC MOLECULES THAT ALLOW THE ORGANISM TO EXTRACT IRON FROM THE
 CC ENVIRONMENT, SPECIALLY UNDER IRON-RESTRICTED CONDITIONS.
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- INDUCTION: BY IRON LIMITATION IN ADDITION TO THE PRESENCE OF ONE
 CC OF THE TWO COGNATE PSEUDOBACTINS Bn8 OR Bm7.
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL; X73598; CAA51995.1; -
CC PIR; S32899; S32899.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC_1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; Signal; Receptor.
FT SIGNAL 1 45 POTENTIAL.
FT CHAIN 46 809 FERRIC-PSEUDOBACTIN BN7/BN8 RECEPTOR.
FT DOMAIN 156 160 POLY-SER.
FT SITE 792 809 TONB C-TERMINAL BOX.
SQ SEQUENCE 809 AA; 88389 MW; 0B339F6E788A8C0D CRC64;

Query Match 18.8%; Score 708.5; DB 1; Length 809;
Best Local Similarity 25.7%; Pred. No. 4.8e-41;
Matches 197; Conservative 154; Mismatches 314; Indels 101; Gaps 23;

QY 8 FRINMTAATVLAAL-----SSSVFAAQT-----DLETVHIKQRSYNAIVTEKNGD 54
DB 94 FDIQGLAIIAGTGLEASRGANASVLSQASASTGALELSAVSISGKAP--GSTTEGTGL 151
QY 55 YSPFVTVGTKIPASLREIPQSVSIITNOQVKORNVDFDQLARKTPGLRVLSNDDG--R 112
DB 152 YTTYSSTSSSTRLNLTPTRETPQSLTWTMRQRLDRLTNLTDALEATPGITVVRDGLGSES 211
QY 113 SVAVRGYSEYSEVNDGLPAQMSINGTLPLNLFADRVVEMRGPSGLDSSGEMGIVNL 172
DB 212 DYSWGRGFAIQYEVGVPTSTR--LDNYSQSWAMFDRVEIVRGAGGLISGMGNPSATINL 270
QY 173 VRKRPTKAPQGHAAAGFGTHQYKAEADVSGSLNSDGSVGRVMAQTVCASPRPAEKNR 232
DB 271 IRKPTAEQAQSIITGEAGNWDRYGTGDFVSGPLTETGNIRGRFVADYKTEKAWIDRYNQ 330
QY 233 HETFAAADWDINDPTVLGAGLYQORHL--APY--NGLPAD--ANNLPSLPQHVFGADWN 289
DB 331 SOLMYGITEFDLSEDTLLTVGFSLYRSIDSDPLRSGLPTFRFTGERTNLKRSLNAAPWS 390
QY 290 KFKMNSHD---VFADLKHVFGNGGKVGKMGVSYDRDADSNGYAFAGSKLGMKTPAGRPGCN 346
DB 391 ---YNDHEQTSFTSIEOQLNGWSGKIELTHAENKFDLNFPA---MGLNPD----- 438
QY 347 TADKACAVGLGTEIKQKALAF-----DASYSRPFRPLGNTANTANEFVIGADYNRFRST 397
DB 439 -----GSLSQLPVRFGTTPRODNLDTATGPPSLFGRHEHLITGWTLSQYREN 487
QY 398 NEQGRITLYARGGLALNEFRSIPQVD-----LIANARKGVRGVSHVATENLDEFGIY 450
DB 488 TFSWGWRYDYAGSPAGAIIDNLFNDGKSAPAFVESGKSSI-----DEDQYAY 537
QY 451 GKSTHPADGLSILGGRLGHYKIES-----GEGKTLHKASKTKFTGTVAGAVYDLDNNN 504
DB 538 LTRRSVTDLSLILGSRUINNRKRTSDRPYGEETEVRNRENGVPIPVAGVGYDLDITW 597
QY 505 SLVLSLSQLYTPQTN--LDADGKLKPRQGNQFQVGYKGSYNMDDRLNARVSFYRMKDKNA 562
DB 598 SLVAYVTKIFNQGAWVTDSEKNPLDPMGEGVGYELGKIGTHLNGKLNSSLAVFKLEQDNL 657
QY 563 AAPLPNNKTYAALGKRVMEGVETEISGAVTPKQWQHAGVSYLHQSIKHTASNRDDGI 622
DB 658 AIWOHDN-----VYSAEQDTSIGIELNGLAELAGWQASAGVSY-----SVTTDADDQRI 708
QY 623 FLMPKHSANLWTTQV---TPELTIGGVNAMSGITSSAGMHA---GGYATFDMAAYR 676
DB 709 NTNLPNSPKTTSRLHGPDLKTIIGGVNWQSKV---GADLHTFSQGSYAVTNLMARYD 766
QY 677 FTFPLKQLQINADINFRHYARVGGANTFNIPGSRBTWTANLRYSF 722
DB 767 INQHSASVNLNNVDFREYYSQSG---LVGVYGTFRNVMTSPKYSF 809

```

RESULT 7

```

YBIL_ECOLI
ID YBIL_ECOLI STANDARD; PRT; 760 AA.
AC P75780;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable tonb-dependent receptor ybil precursor.
GN YBIL OR B0805.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner P.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RT Science 277:1453-1474 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155 (1996).
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY, AND PUTATIVE FUNCTION.
RX MEDLINE=20267815; PubMed=10806384;
RA Molloy M.P., Herbert B.R., Slade M.B., Rabilloud T., Nouwens A.S.,
RA Williams K.L., Gooley A.A.;
RT "Proteomic analysis of the Escherichia coli outer membrane.";
RL Eur. J. Biochem. 267:2871-2881 (2000).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT, THAT PARTICIPATES IN
CC IRON TRANSPORT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).


EMBL; AE000182; AAC73892.1; -  

EMBL; D90717; BAA35471.1; -  

EMBL; D90718; BAA35476.1; -  

EcoGene; EG13317; Ybil.  

InterPro; IPR000531; TonB_boxC.  

Pfam; PF00593; TonB_boxC_1.  

PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.  

PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.  

KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;  

Complete proteome.  

FT SIGNAL 1 33 POTENTIAL.  

FT CHAIN 34 760 PROBABLE TONB-DEPENDENT RECEPTOR YBIL.  

FT SITE 743 760 TONB C-TERMINAL BOX.  

SQ SEQUENCE 760 AA; 81960 MW; 1B66B9C2EF68C7CA CRC64;



Query Match 12.4%; Score 470; DB 1; Length 760;  

Best Local Similarity 22.8%; Pred. No. 1.1e-24;


```


Matches 170; Conservative 109; Mismatches 331; Indels 134; Gaps 23;

QY 65 KLPASLREIPQSVSITITNOQVDRNDVTPDQARKTPGL-RVLSNDGRS---SVYARG 119
 DB 63 KFSRPVADTTRTWTIVSEQVTKQOGATNLDLAKNVPGVGAFAPAGNUNSTTDALYMG 122
 QY 120 YRS-EYINIDGLPAQOSINGTLPNLPAPDRVVMRPGSLFPSSGEMGIVLVKRPPT 178
 DB 123 ADTSNSIYIDGI---RDIGSVSRDFTFTEQVVEVIGKPGSGTDYGRSAPTSIMNISKOFR 178
 QY 179 KAFOGAAAGFGTHQYKKAADVSGSLNSDGRVGRVAGTQVASBRPAKNNRHETFA 238
 DB 179 NSGIDASASISGAMFRKGTLDVNOVIGDTTAVRLNMGKETHDAGDKYKNER----- 232
 QY 239 AADWDINPTVLGAG-----YLYQORHLAPYNGLP-----ADANNKLPBS 277
 DB 233 ---YGVAPSVAPFGLTANLNLNYLHVTCNHTPDGGLPTTGLPGYASAGTAAALHSGK 289
 QY 278 LPOHVVGVADWPKFKRNSHDVFDLKHFGNGYGVKMKRYSRSDADSNY---APAGSKL 334
 DB 290 VTHNHYGTDSDYDDSTTDATATWRFPHDINDNTTIRNTRMSRVKOD---YLMTAINGGAS 347
 QY 335 GCKTPARGG---CNTADKACAVGLGTEIRKQKALAFDASYSRPRLTGTAHEFYIGAD 390
 DB 348 NITQPTSDVNSWTWSRPTANKR-----DVSNNKILTNQTNLTSTYTSIGHDVSTGYE 399
 QY 391 YNRFSTNEQGFRTLYARGL---ALNEFRSIPQVDLIANARKGVSHVATENLDEF 447
 DB 400 FTREPTN-----YGVNPTLPANVNIHPDSIHPCGLTRNGANANGOT-----DTF 446
 QY 448 GIYKSTFHPADGLSIGGRGLGHYKIE-----SGEG-----KTLHK 484
 DB 447 AIIAPFTLTQITRFEINGGIRLNDYHTEVSATACGSGGALTCTGVAKGSPVTTVD 506
 QY 485 ASKTETGY-AGAVYOLNDDNSLYLSLSQLYTPQ-----TNLDAGKILAPRQGS 532
 DB 507 AASGNLMNKAAGLYHLTEGNYVINYAVSQPPGANNPALAQSNGSNSNRITDFKPKRA 566
 QY 533 NOFEVYKGSYMDRLNARVSYFRMKDKNAAPLNNKRTRYAALGKRWMEGVETEISG 592
 DB 567 NISEITKQVLDKLLLRALFRTDIENV---EONDGTYISQYKGRVEGEISVAG 622
 QY 593 AVTPKQIHAGYSYLSQIKTASNSRDDGIFLL--MKPSANLMTTYQVTELLTGGVN 650
 DB 623 NITPAQVIGGYTQOKATIKNGKDVADQSSSLPYTPEHAFTLMSQOATDIDISVGAGAR 682
 QY 651 AMSGITSSAGMAG-----GYATFDMAAYRFPKLTQINADNITNNHYAR 688
 DB 683 YI-----GSMHKGSDGAVGTPTAFTEGYWVADAKLGYRVNRNLDFOLVNLTFTDVSAS 736
 QY 699 VCGANTFNIPGSERTW--TANLRY 720
 DB 737 INKSGYRHPGEPRFTLRLNMAHF 760

RESULT 8
 FOXA_SALTY
 ID FOXA_SALTY STANDARD: PRT: 702 AA.
 AC 056145;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ferritoxamine B receptor precursor.
 GN FOXA OR STM0364.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RP [1]
 RP STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RC MEDLINE=21534948; PubMed=11677609;
 RX McClelland M., Sanderson K.B., Spleeth J., Clifton S.W., Latreille P.,

RA Courtney L., Portolick S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL Nature 413:852-856(2001).
 RN [12]
 RP SEQUENCE OF 1-374 FROM N.A.
 RC STRAIN=SL1344;
 RX MEDLINE=99203118; PubMed=10103258;
 RA Kingsley R.A., Reissbrodt R., Rabech W., Kelsey J.M., Tsolis R.M.,
 RA Everest P., Dougan G., Baemler A.J., Stojiljkovic I., Heffron F.,
 RT "Ferritoxamine-mediated iron(III) utilization by Salmonella
 enterica.";
 RL Appl. Environ. Microbiol. 65:1610-1618(1999).
 RN [13]
 RP SEQUENCE OF 1-244 FROM N.A.
 RC STRAIN=ATCC 14028;
 RX MEDLINE=95370139; PubMed=7642488;
 RA Tsolis R.M., Baemler A.J., Stojiljkovic I., Heffron F.,
 RT "Fur regulon of Salmonella typhimurium: identification of new iron-
 regulated genes.";
 RL J. Bacteriol. 177:4628-4637(1995).
 CC - FUNCTION: FERRITOXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE
 TONB PROTEIN (BY SIMILARITY). MAY PLAY A ROLE IN INTESTINAL
 COLONIZATION.
 CC - SUBCELLULAR LOCATION: Outer membrane.
 CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC - This SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AE008712; AAL19318.1; -;
 DR EMBL, AF060876; AAC15464.1; -;
 DR EMBL, U62282; AAB04552.1; -;
 DR HSSP, P06971; 1BY5.
 DR StyGene; SG10646; foxA.
 DR InterPro; IPR000531; TONB boxC.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB box; Complete proteome.
 FT SIGNAL 1 30
 FT CHAIN 1 702
 FT CONFLICT 246 702 FERRITOXAMINE B RECEPTOR.
 FT CONFLICT 246 249 MAYP -> SPVL (IN REF. 2).
 FT CONFLICT 255 255 G -> R (IN REF. 2).
 FT CONFLICT 279 279 F -> S (IN REF. 2).
 FT CONFLICT 352 352 A -> P (IN REF. 2).
 FT CONFLICT 368 374 HRLIVGI -> SSSAGGD (IN REF. 2).
 FT FT
 SQ SEQUENCE 702 AA; 77685 MW; 7E2CBA405B09E860D CRC64;

Query Match 11.2%; Score 423; DB 1; Length 702;
 Best Local Similarity 23.1%; Pred. No. 1.6e-21;
 Matches 176; Conservative 119; Mismatches 360; Indels 106; Gaps 25;

QY 5 MSVFRINMTAAATVLAALSS---VFPAQTADLETYHIKQGRSYNAIVTEKNGDYSFAVT 61
 DB 5 MEMFATRLMALLGATIGATPELPFAQETTKNDV-----IVT-----SPVGS 47
 QY 62 VGTKLPASLREIPQSVSITITNOQVDRNDVTPDQARKTPGLRVLSNDGRS---VYA 117
 DB 48 GATKLTATPIETPQSVSITITROQFEQATSVRAQAVSYTPG--VYSNQGASNRFDYIVL 105
 QY 118 RGYEYSEYN---IDGLP-AQMQSINGTLPNLPAPDRVVMRPGSLFPSSGEMGIVLV 172
 DB 106 RGRSDSLNLYVLDGLKMGMDTNSHSLVVDPWFLIEDIVRGVAPSVLYGRSSPGIVLV 165


```

Db      227 REERYAIASLLMQPDENSTLLRANLQKDPSSGYY-----HSNVPAADGSIYQKLS 277
Qy      280 OHFVADVADNWKFGNSHDVFAADLKHYFGNGGYGVGMRYSDRDADNSVYAFAGSKLGKTP 339
Db      278 RGFDE-----SHNVFKRQOQIYSEFSKFPDDWSFRQ-NASVTHSNQLEQVYQ 329
Qy      340 AGPGCNTADKACAVGLGTEIKKALAPDASVSRPRLGNTANEPVIGADYVFRS--- 396
Db      330 GGNSDRTLMNRYYS---GEDSLNPAVDNQLADRLTAHVHKKVLLGVDPQKFRNNLR 386
Qy      397 -----TNEGRTTLVARGGLANLFRSIPQVDLIANARKGVRSHTVATENL 444
Db      387 SDSAVATPLNPPYGVSGSSTLYS-----DYLRT-----TPGINTSYLSRRY 427
Qy      445 DERGIYKSTFHPADGLSLIGGRGLGHYKIE-----SGEGTLHKASVTKFTGYAGAV 497
Db      428 EOGSVTLQDEM-TLDWMHNLTSGRYDRMKTENINNTANSTDETDNHA-----GRASL 481
Qy      498 YDLNDNNSLYLSQLYTPQTNLDADGKLKPRQNGQFEVG---YKGSYMDRLNARVSFY 555
Db      482 YSPDSGISPYVSQAITPBLPDAQQKLLKPMTSQYEVGIIYPPGSGTSLVSALYDL 541
Qy      556 RKMD-KVAAAPLPNNKKTYYALGKRVMEGVETEISGAVTPKKQIHAGYSYLSQIKTA 614
Db      542 TONDVANRAVP-----ATYYVPAGVYNSQGLELARSQISDRLSVIAGTYNRVFXKA 595
Qy      615 SNSRDGIFFLMPKHSANLMTTYQVTPPELLIGGVNAMSIG-----TSSAGMHAGY 666
Db      596 IGDNDNTVLABSNWASLMAQTEA-----GYGVNGAGIRYIKQWADANTLVPST 649
Qy      667 AAFDAAVY---RFTPKLK---LQINADNIFNRHYARVAGANTFNI PGSERTWTANLRY 720
Db      650 TLGDASVRADLGWASLKGAFVQLVNNINADKKVYAA-CYSTSYCWAGERSVQATVGY 708
Qy      721 SF 722
Db      709 DF 710

RESULT 10
FHUA_ECOLI
ID      FHUA_ECOLI STANDARD; PRT; 747 AA.
AC      P06971; P75665; P71280;
DT      01-APR-1988 (Rel. 07, Created)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Ferrichrome-iron receptor precursor (ferric hydroxamate uptake)
GN      FHUA OR TONA OR B0150.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
CC      NCBI_taxid=562;
OK      (1)
RP      SEQUENCE FROM N.A., AND SEQUENCE OF 34-47.
RX      STRAIN=K12;
RC      MEDLINE=86085668; PubMed=3079747;
RA      Coulton J.W., Mason P., Cameron D.R., Carmel G., Uean R., Rode H.N.;
RT      "Protein fusions of beta-galactosidase to the ferrichrome-iron
RL      receptor of Escherichia coli K-12.";
RL      J. Bacteriol. 165:181-192(1986).
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / W3110;
RX      MEDLINE=94261430; PubMed=8202364;
RA      Fujita N., Mori H., Yura T., Ishihama A.;
RT      "Systematic sequencing of the Escherichia coli genome: analysis of
RL      the 2.4-4.1 min (110,917-193,643 bp) region.";
RL      Nucleic Acids Res. 22:1637-1639(1994).
[3]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MG1655;

```

```

RX      MEDLINE=97426617; PubMed=9278503;
RA      Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
[4]
RP      SEQUENCE OF 482-647 FROM N.A.
RX      Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA      Davis K., Federspiel N., Hyman R., Kaitman S., Komp C., Kurdi O.,
RA      Laskari D., Lew H., Lin D., Namach A., Oelner P., Roberts D.,
RA      Davis R.W.;
RT      Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
[5]
RP      SEQUENCE OF 723-747 FROM N.A.
RX      MEDLINE=88038363; PubMed=2823072;
RA      Burkhardt R., Braun V.;
RT      "Nucleotide sequence of the fhuc and fhud genes involved in iron
RT      (III) hydroxamate transport: domains in Fhuc homologous to
RT      ATP-binding proteins.";
RL      Mol. Gen. Genet. 209:49-55(1987).
[6]
RP      REVIEW.
RX      MEDLINE=94265918; PubMed=7515827;
RA      Braun V., Killman H., Benz R.;
RT      "Energy-coupled transport through the outer membrane of Escherichia
RT      coli small deletions in the gating loop convert the fhua transport
RT      protein into a diffusion channel.";
RL      FEBS Lett. 346:59-64(1994).
[7]
RP      ION CHANNEL.
RX      MEDLINE=96203106; PubMed=8617231;
RA      Bonhivers M., Ghazi A., Boulanger P., Letellier L.;
RT      "Fhuda, a transporter of the Escherichia coli outer membrane, is
RT      converted into a channel upon binding of bacteriophage T5.";
RL      EMBO J. 15:1850-1856(1996).
[8]
RP      X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 34-747.
RX      MEDLINE=99074366; PubMed=9856937;
RA      Ferguson A.D., Hofmann E., Coulton J.W., Diederichs K., Welte W.;
RT      "Siderophore-mediated iron transport: crystal structure of Fhua with
RT      bound lipopolysaccharide.";
RL      Science 282:2215-2220(1998).
[9]
RP      X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 53-747.
RX      MEDLINE=99081293; PubMed=9865695;
RA      Locher K.P., Rees B., Koebnik R., Mitschler A., Moulinier L.,
RA      Rosenbusch J.P., Moras D.;
RT      "Transmembrane signaling across the ligand-gated Fhua receptor:
RT      crystal structures of free and ferrichrome-bound states reveal
RT      allosteric changes.";
RL      Cell 95:771-778(1998).
[10]
RP      X-RAY CRYSTALLOGRAPHY (2.95 ANGSTROMS) OF 34-747.
RX      MEDLINE=20306978; PubMed=10850805;
RA      Ferguson A.D., Braun V., Fiedler H.-P., Coulton J.W., Diederichs K.,
RA      Welte W.;
RT      "Crystal structure of the antibiotic albomycin in complex with the
RT      outer membrane transporter Fhua.";
RL      Protein Sci. 9:956-963(2000).
-1- FUNCTION: THIS RECEPTOR BINDS THE FERRICHOME-IRON LIGAND. IT
INTERACTS WITH THE TONB PROTEIN, WHICH IS RESPONSIBLE FOR ENERGY
COUPLING OF THE FERRICHOME-PROMOTED IRON TRANSPORT SYSTEM. ACTS
AS A RECEPTOR FOR BACTERIOPHAGE T5 AS WELL AS T1, PH180 AND
COLICIN M. BINDING OF T5 TRIGGERS THE OPENING OF A HIGH
CONDUCTANCE ION CHANNEL. CAN ALSO TRANSPORT THE ANTIBIOTIC
ALBOMYCIN.
-1- SUBUNIT: MONOMER
-1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
-1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M12486; AAB61768.1; -;
 DR EMBL; D25662; BAA05598.1; -;
 DR EMBL; AE000124; AAC73261.1; -;
 DR EMBL; U70214; AAB08580.1; -;
 DR EMBL; X05810; CAA29253.1; -;
 DR PIR; A25196; QRECFE.
 DR PIR; S06358; S06358.
 DR PIR; S45219; S45219.
 DR PDB; 1BY3; 13-JAN-99.
 DR PDB; 1BY5; 13-JAN-99.
 DR PDB; 1FCP; 13-JAN-99.
 DR PDB; 2FCP; 13-JAN-99.
 DR PDB; 1OQJ; 05-JUN-00.
 DR PDB; 1OQC; 05-JUN-00.
 DR EcoGene; EG10302; fluA.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC_1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Iron transport; Transpor; TonB box; Signal;
 KW Phage recognition; Receptor; Transmembrane; 3D-structure;
 KW Complete proteome.
 FT SIGNAL 1 33
 FT CHAIN 34 747
 FT DOMAIN 34 747 FERRICHRONE-IRON RECEPTOR.
 FT TRANSMEM 193 201 PERIPLASMIC.
 FT DOMAIN 202 206 EXTRACELLULAR.
 FT TRANSMEM 207 215
 FT DOMAIN 216 222 PERIPLASMIC.
 FT TRANSMEM 223 231
 FT DOMAIN 232 245 EXTRACELLULAR.
 FT TRANSMEM 246 255 PERIPLASMIC.
 FT DOMAIN 256 259
 FT TRANSMEM 260 268 EXTRACELLULAR.
 FT DOMAIN 269 312
 FT TRANSMEM 313 321
 FT DOMAIN 322 326 PERIPLASMIC.
 FT TRANSMEM 327 335 EXTRACELLULAR.
 FT DOMAIN 336 387
 FT TRANSMEM 388 396
 FT DOMAIN 397 404 PERIPLASMIC.
 FT TRANSMEM 405 413 EXTRACELLULAR.
 FT DOMAIN 414 464
 FT TRANSMEM 465 473 PERIPLASMIC.
 FT DOMAIN 474 477
 FT TRANSMEM 478 486 EXTRACELLULAR.
 FT DOMAIN 487 508
 FT TRANSMEM 509 517
 FT DOMAIN 518 522
 FT TRANSMEM 523 531 PERIPLASMIC.
 FT DOMAIN 532 551
 FT TRANSMEM 552 560 EXTRACELLULAR.
 FT DOMAIN 561 565
 FT TRANSMEM 566 574 PERIPLASMIC.
 FT DOMAIN 575 601
 FT TRANSMEM 602 610 EXTRACELLULAR.
 FT DOMAIN 611 613
 FT TRANSMEM 614 622 PERIPLASMIC.
 FT DOMAIN 623 645
 FT TRANSMEM 646 654
 FT DOMAIN 655 661
 FT TRANSMEM 662 670 EXTRACELLULAR.
 FT DOMAIN 671 689
 FT TRANSMEM 690 698
 FT DOMAIN 699 705
 FT TRANSMEM 706 714 PERIPLASMIC.

FT DOMAIN 715 737
 FT TRANSEM 738 746
 FT DOMAIN 747 747 PERIPLASMIC.
 FT SITE 40 47 TONB BOX.
 FT SITE 730 747 TONB C-TERMINAL BOX.
 FT BINDING 114 114 FERRICHRONE.
 FT BINDING 133 133 FERRICHRONE.
 FT BINDING 148 149 FERRICHRONE.
 FT BINDING 277 277 FERRICHRONE.
 FT BINDING 299 279 FERRICHRONE.
 FT BINDING 346 346 FERRICHRONE.
 FT BINDING 348 348 FERRICHRONE.
 FT BINDING 424 424 FERRICHRONE.
 FT BINDING 735 735 FERRICHRONE.
 FT DISULFID 351 362
 FT DISULFID 725 731
 FT CONFLICT 609 610
 FT CONFLICT 737 737
 SQ SEQUENCE 747 AA; 82182 MW; 1C2B251D1990E444 CRC64;
 AA -> RP (IN REF. 2).
 R -> P (IN REF. 5).
 Query Match 10.3%; Score 387.5; DB 1; Length 747;
 Best Local Similarity 22.0%; Pred. No. 4.9e-19;
 Matches 175; Conservative 122; Mismatches 362; Indels 135; Gaps 30;
 QY 6 SVFRINMTAATVLAALSSSVFAAQADLTETVHIKQSRNAIVTEKNGDYSSPAVTVTGK 65
 DB 12 SLRKIAVVAVTAVSGMSVYAAQAVEPKEDITITVTAAPQESAWGPAATIAAROSATGK 71
 QY 66 IPASLRIPQSVSIITNQKDRNVDTFDQLARKTPGLRVLSDDDGRSSVY----- 116
 DB 72 TDTPIQKVPQISVVTAEEMALHPKSVKEALSYPGVSV--GTRGASNTDHLIIRGFA 129
 QY 117 ARGYEYSEYNTDGLPAQMSINGTLPNLPADRVEMRGPSGLFDSSGEMGGINLVKVR 176
 DB 130 AEGSQNNY-LNGELKQGNFYNDVIDPYMLERAERINRGVSVLYGKSSPGGLLNWVSKR 188
 QY 177 PT-----KAFQCHAAAGFTGHQYKAEADVSGSLNSDG--SVRGRVMAQTVCASPRPAEK 230
 DB 189 FTTEPLKEVQPKA-----GTDLSLFTQTFDFDSLDSDDDGVYSYRLTGLARSANAQ----QKG 240
 QY 231 NRHETFAAADWDINPD-----TVLGAGLYQORHLAPYNGLPADAN-NKLPS---LPQH 281
 DB 241 SEEQRYAIAPAFTWRPDDKTNFTFL--SYFQNEPETGYGHLPEKGVTEPLNGKLPD 298
 QY 282 VFGADWNKFKQNSHDVPADLKHVFGNGGYGKGVKRYSDRDADSNYPAGSKLGMKTPAG 341
 DB 299 FNEGAKNNTYSRNEKMGVGYSPDHEFND-----TFTVR---QNLRAFNKTSQNSVYG 347
 QY 342 RPGCN--TADDKACAV-----GLGTEIKOKALAFDASYSRPFRLGNTANEF 385
 DB 348 YGVCSDPANAYSQCAALAPADKGHYLARKYVVDDEKLQNFSDVTQLOSKFATGDIDHTL 407
 QY 386 VIGADYNRFRSTNEQRTTLYARGGLALNEPRSIPOVDLIANARKVGRGYSHTVATENLD 445
 DB 408 LTGVDFWEMR--NDINAWFGYDDSVPLLLNLYNPV-NTDFDNADKPANSGPYRLNKQ-K 463
 QY 446 EFGIY-----GKSTFHPADGLSLIGGRLGHYKIESGEGKTLHKASKTKF 490
 DB 464 QTGVYVQDAQWKVLVTLGGRYDWAQDES LNRVAG-----TTDKRDDKQF 509
 QY 491 TGYAGAVYDNDNNSLYLSLSQLYTPQTNLDADGKLLKPROGNQFVGYKGSYMDRLNA 550
 DB 510 TWRGVNYLFDNGVTPTFSYSESPSSQVQKQGNIFAPSKGQYGVGVKVPDRPLV 569
 QY 551 RVSFYRMKDQVAAAAPLNPNNKKTTRYAALGKRV-MEGVETEISGAVTPKWQIHAGYSYLHS 609
 DB 570 TGAVYNTKTN--NLMAADPGSFFSVEGGEIRARGVEIEAKAALSASVNVVGSYTYTDA 626
 QY 610 QIKTASNSRDDGIFLLMPKHSANLWTTVQVTPPELTIGGVNMMSGITSSAG----- 660
 DB 627 EYTTDTTVYKGN-TPAQVPKHMASLWADY-----TFDDG--PLSGLTLGTGRVTGSSYG 677
 QY 661 -----MHAGGYATFDMAAYRFT-----PKLKLQINADNIFNRHYARVVGANTFN-IP 708


```

Db      678 DPANSTKVSSTYVVDALVRIDLARVGMAGSNVALHVN--NLFREYVASC--ENTYGCPEW 733
QY      709 GSERTWTANLRYSF 722
Db      734 GABROVAVATATRRF 747

RESULT 11
FCT_ERMCH STANDARD; PRT; 735 AA.
ID_FCT_ERMCH
AC 04762:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Ferrichrysoabactin receptor precursor.
GN FCT.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937;
RX MEDLINE=96165286; PubMed=8576065;
RA Sauvage C., Franza T., Expert D.;
RT "Analysis of the Erwinia chrysanthemi ferrichrysoabactin receptor
RT gene: resemblance to the Escherichia coli fepA-fes bidirectional
RT promoter region and homology with hydroxamate receptors.";
RL J. Bacteriol. 178:1227-1231(1996).
CC - FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING
CC CHRYSOABACTIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS THE
CC BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC - SUBCELLULAR LOCATION: Outer membrane.
CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X87967; CAA61205.1; -.
DR HSSP; P06971; 2PCP.
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE_NEG.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
DR Transport; TonB box.
FT SIGNAL 1 38
FT CHAIN 39 735
FT SITE 45 52
FT SITE 718 735
FT DOMAIN 39 45
FT TRANSMEM 46 54
FT TRANSMEM 55 82
FT TRANSMEM 83 91
FT TRANSMEM 92 107
FT TRANSMEM 108 116
FT TRANSMEM 117 124
FT TRANSMEM 125 133
FT TRANSMEM 134 169
FT TRANSMEM 170 178
FT TRANSMEM 179 179
FT TRANSMEM 180 188
FT TRANSMEM 189 194
FT TRANSMEM 195 203
FT TRANSMEM 204 222
FT TRANSMEM 223 231
FT TRANSMEM 232 261
FT DOMAIN 261
FT PERIPLASMIC (POTENTIAL).

```

```

FT TRANSMEM 262 270
FT DOMAIN 271 286
FT TRANSMEM 287 295
FT TRANSMEM 296 312
FT TRANSMEM 313 321
FT TRANSMEM 322 328
FT TRANSMEM 329 337
FT TRANSMEM 338 343
FT TRANSMEM 344 352
FT TRANSMEM 353 378
FT TRANSMEM 379 387
FT TRANSMEM 388 392
FT TRANSMEM 393 401
FT TRANSMEM 402 403
FT TRANSMEM 404 412
FT TRANSMEM 413 424
FT TRANSMEM 425 433
FT TRANSMEM 434 448
FT TRANSMEM 449 457
FT TRANSMEM 458 461
FT TRANSMEM 462 470
FT TRANSMEM 471 495
FT TRANSMEM 496 504
FT TRANSMEM 505 509
FT TRANSMEM 510 518
FT TRANSMEM 519 538
FT TRANSMEM 539 547
FT TRANSMEM 548 552
FT TRANSMEM 553 561
FT TRANSMEM 562 568
FT TRANSMEM 569 577
FT TRANSMEM 578 599
FT TRANSMEM 600 608
FT TRANSMEM 609 649
FT TRANSMEM 650 658
FT TRANSMEM 659 674
FT TRANSMEM 675 683
FT TRANSMEM 684 695
FT TRANSMEM 696 704
FT TRANSMEM 705 706
FT TRANSMEM 707 715
FT TRANSMEM 716 726
FT TRANSMEM 727 735
SQ SEQUENCE 735 AA; 81055 MW; 3C37073BE4538DC3C CRC64;

Query Match 9.9%; Score 374; DB 1; Length 735;
Best Local Similarity 23.8%; Pred. No. 4,1e-18;
Matches 184; Conservative 119; Mismatches 341; Indels 130; Gaps 31;

QY 14 AATVLAALSSVFAAQTADLETVHIKQGRS-----YNAIVTEKNDYSSFAVVGTKIP 67
Db 27 ASTLLMAHAHQ--AAETGADTMIVSANAGESVTAPLKIIVAKESMS-----GTKTS 76
QY 68 ASLREIPQSVSIITNQVQRNVDTFDQLARKTPGLRVLNSDDGRSS---VYAR-GVEY 122
Db 77 TPLIKTPQSVTVVTAATKMDAQAVSSVSHALNYSQG--VVTNRGSSNNRDEVIARVRVY 134
QY 123 SEYNIDGLPAQMSINGTL--PNTLAFDRAVEYMRGPGSGIFDSSGEMGCIYVLVRRPPTA 180
Db 135 APKFIDGLSYGSGSGSTIGKNPMLERVENHGPAVLYQGVNPGGISTSKRPTA- 193
QY 181 FQGHAAAGGTHKQYKAE--DVSGSLNSDGSVGRGVMAQTGASPRAPKKNRHEFTFA 238
Db 194 -ETIKRVQFSAGNQHGEAAPFGGALNDKTLVRL--DGLASTGHEVVKDSKDERIAV 250
QY 239 AAD--WDINPDVLGAGLYQQRHLAPYNGLPADANNKLPSPQHVFGA----- 286
Db 251 APSLTLPLPDPDTSFTLLTSYQNDPRAGY-----RNFLPKIGTVVAGAGVTPYDLNVS 303
QY 287 --DNKFKRNSHDVPAADLKHVYGNNGYGVGMKRYSDRDADSNYAFAGSLGKMTAGRG 344
Db 304 DPNYQSKREGSGISGVNLDHSFNDVFSFOQNVRYTOLREKYKY-----LVYTKNADAPA 357

```


QY 345 CNTA-----DDKACAVGLGTEIKOKALAFDASYSRPRFLGNTANEFVIGADYNRPSTNEQ 400
 Db 358 TDTTLIRRPQK-----BENEISEFAIDNQLKATFATGVSNNHTVLSGLDYKWL----- 404
 QY 401 GRTTLYARGGLALNEFRSPQVDLIANARKVGRVGSHTVATE-----NLDEFGIYGKS 453
 Db 405 ---TLEKOWLDRNDYSFNWNPYV-----SVNDSMLTELSTNERNKLNQGVYLOD 456
 QY 454 TPHPADGLSLIGGRGLGHYKIESGE--GKTLHKASKTFTGYAGAVYDLNDNNSLYLSLS 511
 Db 457 QLE-WNQNLLLSGRHDSVRDKQDYAADTTTTERNDGKFTRAVRLLYAFDNGISPVVSYS 515
 QY 512 QLYTPTNLDADG-KLLKPRQNGQFVGYKSGYMDRLNARVFRMKDKNAAPLNPN 570
 Db 516 TSFEPNLDGAPGTFAPKPTTGEQEVGKFPQKGSNTLLTVSLFDITQKN-----ITSYN 571
 QY 571 KTRXY-AALGKRVMEGVETEISGAVTPKQIHHAGYSYLH-----SOIKTASNSRDD 620
 Db 572 SVTRNEQIGKVKSGVETEAHTQLTPEISLMAAYSITDAVTKESYASQVNKAPSS----- 628
 QY 621 GIELMPKHSANLWTTYQV-----TPELTIGGGVNAMSGITSSAGMHAGGYATFDMAAYR 676
 Db 629 -----IPRAHASAWGSYSPHNGPLKGLTGTVRYI-GSTTAITRASRYRLYPVRRHGR 682
 QY 677 F-----TPKLK--LQINADNIFNRHYARVGGANTFNIPGSERTWTANLRYSF 722
 Db 683 YELGSLASQLKGAALVOLNNLTDRHYVASCQG-DTACFYSGGRTVATVSYSW 735

RESULT 12
 FYUA_YEREN
 ID FYUA_YEREN STANDARD; PRT; 673 AA.
 AC P46360;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Pesticin receptor precursor (IRPC) (IPR65).
 GN FYUA.
 OS Versinia enterocolitica.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Versinia.
 OX NCBI_TaxID=630;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;
 RX MEDLINE=95075311; PubMed=7984105;
 RA Rakin A., Saken E., Harmsen D., Heesemann J.;
 RT "The pesticin receptor of Versinia enterocolitica: a novel virulence
 factor with dual function.";
 RL Mol. Microbiol. 13:253-263 (1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1223-75-1, 8081, YE737, and 1209-79;
 RA Rakin A., Heesemann J.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
 CC SIDEROPHORE VERSINIABACTIN.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z29675; CAA82773.1; -
 DR EMBL; Z35485; CAA84620.1; -
 DR EMBL; Z35486; CAA84621.1; -
 DR EMBL; Z35487; CAA84622.1; -
 DR EMBL; Z35496; CAA84630.1; -

DR HSP; P05825; IPEP.
 DR InterPro; IPR00531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
 DR PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB_box. 22
 FT SIGNAL 1 POTENTIAL.
 FT CHAIN 23 673 PESTICIN RECEPTOR.
 FT SITE 30 37 TONB BOX.
 FT SITE 657 673 TONB C-TERMINAL BOX.
 FT VARIANT 362 362 N -> D (IN STRAINS 8081, YE737 AND 1209-
 FT SEQUENCE 673 AA; 73682 MW; 1C20E0352FAD4DCC CRC64;
 SQ
 Query Match 8.0%; Score 303; DB 1; Length 673;
 Best Local Similarity 20.9%; Pred. No. 2.7e-13;
 Matches 159; Conservative 98; Mismatches 298; Indels 204; Gaps 30;
 QY 54 DYSSFAVTGKIPASUREIPQSVSIITNQOVKORNVDTFDQLARKTPGLRVLSNDDGR- 112
 Db 28 DESTLEVTASKQSSRSASANNVSVVSAPELSDAGVTASDKLPVLPLGLNI--ENSNGM 85
 QY 113 --SSVYARGYEYSB--YN-----IDGLPAQMQSINGTLPNLFAPDRVENVMRGSGLPDS 162
 Db 86 LFTSLIRGVSQAQDFYNPAVTLVYDGVF---QLSTNTIQLTVDQSVELLRGPGQGLYVG 142
 QY 163 SGEAGIYNLVKRPKTKAFQGHAAAGFGTHKQYKAEADVSGSLNSDGSVGRVNAQTVGA 222
 Db 143 KSAQGGIINIVTQPDSTPRGYIEGGVSSRDSYRSKFNLSGPI-QDGLLYGSV----- 194
 QY 223 SPRPAEKNNRHETFYAAADWD-INPDTVLGAGYLYQORHLAPYNGLPADANKPLPSLPQH 281
 Db 195 -----TLRQVDDGDGMINPAT----- 210
 QY 282 VFGADWNKFKWNSHDVPADLKHYFGNGG---YGVKVMYSDRDADSNVAFAGSKLGMK 337
 Db 211 ---GSD-----DL-----GGTRASIGNVKRLRAPDDQPWEMGFAASR----- 244
 QY 338 TPAGRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPRFLGNTANEFVIG- -DY 391
 Db 245 -----ECRATQDA-YVG-WNDIKGRKLSLSDGSPDPYMRRCCTDSQTLSCGYTTDDWVF 296
 QY 392 NRPSTNEQGRITLYARGGLALNEFRSIPO-----VDLI-----AN 427
 Db 297 NLISAWQQOQHYSTRFPSSGLIVN---MPQRWNQDVQELRAATLGDARTVDMVFLYRQN 352
 QY 428 ARKGVRG-----YSHTVATENLDEFGIYKSTFHPADGLSLIGGRGLGHYKIESG-E 478
 Db 353 TREKLNSAYNMPYPLSSSTGYTAEITLAAYSDLTWHLTDRFDIGGGVRFSDHKSSTQYH 412
 QY 479 GKTL-----HKASKTKFTGYAGAVYDLNDNNSLYLSLSQLYTPQ-----TNLDADGKL 526
 Db 413 GMLGNPFGDQKSNDDQVLGQLSAGYMLTDDWRVYTRIAQGYKPSGVNIPTAGLDAKP 472
 QY 527 LKPRQNGQFVGYKSGYMDRLNARVFRMKDKNAAPLNPNKKTRYAALGKRVMEGV 586
 Db 473 FVAESKSYNELGTREYETADVTLOAFTYTHTKDMLQYS--GPVGMQTLNSA-GRADATGV 529
 QY 587 ETEISGAVTPKQIHHAGYSYLHSGI-----KTASNSRDDGIF-----LL 625
 Db 530 ELEAKRFPAGSWDINGNIRSRSEFTNDSELYHGNRVFPVPRYAGGSSVNVIDTRYGAL 589
 QY 626 MPKHSANLWTTYQVTPELTIGGGVNAMSGITSSAGMHAGGYATFDMAAYRFTPKLQI 685
 Db 590 MRLAVNL-----VGPYFDGDN-----QLRQGTATLTDSSIGWQATERINISV 633
 QY 686 NADNIFNRHY--YARVGGANTFNIPGSERTWTANLRYSF 722
 Db 634 HVDNLFDRRYRTYGYMGSSAVAQVNMGRVTGINTRIDF 672

RESULT 13

EYUA_YERPE STANDARD; PRT; 673 AA.
 AC P46359;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Pesticin receptor precursor (IRPC).
 GN EYUA OR PSN OR YPO1906.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OC NCBI_Taxid=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6/69;
 RA Rakin A., Heesemann J.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KIM6;
 RX MEDLINE=95204350; PubMed=7896707;
 RA Fetherston J.D., Lillard J.W., Perry R.D.;
 RT "Analysis of the pesticin receptor from Yersinia pestis: role in
 RT iron-deficient growth and possible regulation by its siderophore."
 RL J. Bacteriol. 177:1824-1833(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=6/69;
 RA Buchrieser C., Rusnok C., Couve E., Frangoul L., Billault A.,
 RA Kunst F., Carniel E., Glaser P.;
 RT "DNA sequence of the 102 kbases unstable region of Yersinia pestis."
 RT Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parthill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Bastam D., Bentley S.D., Brookes K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Fellwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 CC - FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE
 CC SIDEROPHORE YERSINIBACTIN.
 CC - SUBCELLULAR LOCATION: Outer membrane.
 CC - INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH
 CC THE PUR PROTEIN.
 CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; Z55104; CAA84487.1; -;
 DR EMBL; U09530; AAA69906.1; -;
 DR EMBL; AL031866; CAA21395.1; -;
 DR EMBL; A1414150; CAC90722.1; -;
 DR HSPB; P05825; IFEF.
 DR InterPro; IPR000531; TONB_BOX.
 DR Pfam; PF00593; TONB_BOX; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
 KW Transport; TonB box; Complete proteome.
 FT SIGNAL 1 22 POTENTIAL.
 FT CHAIN 23 673 PESTICIN RECEPTOR.

FT SITE 30 37 TONB BOX.
 FT SITE 657 673 TONB C-TERMINAL BOX.
 FT CONFLICT 484 484 G -> D (IN REF. 1).
 FT CONFLICT 514 514 R -> G (IN REF. 1).
 SQ SEQUENCE 673 AA; 73782 MW; 9C39E6010BCEB2C CRC64;
 Query Match 7.9%; Score 300; DB 1; Length 673;
 Best Local Similarity 20.9%; Pred. No. 4,4e-13;
 Matches 159; Conservative 97; Mismatches 299; Indels 204; Gaps 30;
 QY 54 DVSSFAVYVGTIKPISLRITPOSVSIITNQYKDRVDFPDQARTPLRLVLSNDGR 112
 DB DESTLVVTASKSSRSASANNVSTVASPELSDACVATSDKLPRLVPLINT--ENSGNM 85
 QY 113 --SSVYARGVYSE--YN-----IDGLPAQMSINGTLPLNFAFDRVEMRGPGLFDS 162
 DB LFTSLIRGVSSAQDPYNAVTLVYDGV-QLSTNTIQLTVDVSVELIRGPSTLYG 142
 QY 163 SGEMGIIVLVKRPPTKAFQGAAGFTHKQYKADVSGLSNDSVGRVMAQTGVA 222
 DB KSAQGIINIIVTQPDSTRGYIEGVSRSDSYRSKFNLSGPI-QDGLYGSV----- 194
 QY 223 SRPRAKNNRHETFYAADMW-INPDTVAGVLYQGRHLAPNGLPADANNKPSLPQH 281
 DB 195 -----TLRQVDGDMINPAT----- 210
 QY 282 VFGADNFKKFNKNSHDFADLKHVFGNG-----YGVKGRYSRDRADSNVAFASKLGK 337
 DB 211 ---GSD-----DL-----GSTRASIGVVKRLAPDDOPWEMGRFASR---- 244
 QY 338 TAGRPGCNTADKACAVAGTEIKOKALAPASYSRPRLTANTANEVIGA-----DY 391
 DB 245 -----ECTRATQDA-YVG-WNDIKGRKLSISGSDPDPFRCTDSGLTKYTDDWF 296
 QY 392 NRPRTNEGRRTTLVARGGLANERSIPO-----VDLI-----AN 427
 DB 297 NISAMQOQHYSRTFPGSLIVN---MPQRNVDQVQLRAATLGADRTVDVFGLYRON 352
 QY 428 ARKGVG-----YSHVAVATENIDFQYKSPFHPADGSLIGGGLGHYKIESG-E 478
 DB 353 TREKLSAYDMTMYPLYSTGTTTETLAYSDLTMLTDRPDIGGVAFSHDKSTQYN 412
 QY 479 GXTL-----HKASKTKTYGAAVYDLNDDNSLYLSLQUTPQ-----TNLDADGKL 526
 DB 413 GSWLGNPFQDQKSDNDQVLGQLSGVYMLTDMRVYTRAAQGVKPSGVNIIVPAGADAP 472
 QY 527 LKPRGNGPEVGVKGSYMDRLNARVSPYRMDKNAAPLANPNKKTRYAALGKRVMEGV 586
 DB 473 FVAEKSYNELGTRYETADVTLQAATFTYHTRDMOLYS--GEVRMQTLISNA-GKADATGV 529
 QY 587 ETEISGAVTPKQOIHAGVSYLSQI-----KTASNRDGIIF-----LL 625
 DB 530 ELBAKRPAFGMSWDINGNVIRSEFTNDELYHGNRPVFPVRYGASSVYNGVITDRYGL 589
 QY 626 MKPSHNLMTTQVTPPELLTIGGVNAMSGITSSAGAHAGVATFPDMAAYRFTPKLKLOI 685
 DB 590 MRLAVNL-----VGHYVDGN-----QLRQSTYATLDSLSLQWATERNNISV 633
 QY 686 NADNIFNRHY--YARVGANTNIPGSERTYANLRYSF 722
 DB 634 YVDNLFDRRYRTGYWNGSSAVAQVVMGRTVGINRRIDF 672
 RESULT 14
 CIRA_ECOLI STANDARD; PRT; 663 AA.
 AC P17315;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Colicin I receptor precursor.
 GN CIRA OR CIR OR FEUA OR B2155.
 OS Escherichia coli.

DB 620 ANQVTVKDVLRAGVNL 636

RESULT 15

PFEA_PSEAE STANDARD; PRT; 746 AA.

AC 005058; 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ferric enterobactin receptor precursor.

GN PFEA OR PA2688.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K407;

RX MEDLINE=93123148; PubMed=8419284;

RA Dean C.R., Poole K.;

RT "Cloning and characterization of the ferric enterobactin receptor gene (flea) of Pseudomonas aeruginosa.";

RT J. Bacteriol. 175:317-324 (1993).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M., Gader R.L., Goltz U., Tolentino E., Westbrock-Wadman S., Yan Y., Brady L.L., Coulter S.N., Folger K.R., Kas A., Laibig K., Lim R.M., Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T., Reiter J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen.";

RT Nature 406:959-964 (2000).

RL Nature 406:959-964 (2000).

CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC ENTEROBACTIN.

CC -1- SUBCELLULAR LOCATION: Outer membrane.

CC -1- INDUCTION: BY IRON AND ENTEROBACTIN.

CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL; M98033; AAA25928.1; -

DR EMBL; AE004697; AAG06076.1; -

DR PIR; A40636; A40636.

DR HSSP; P05825; 1FEP.

DR InterPro; IPR000531; TONB_boxC.

DR Pfam; PF00593; TONB_boxC.1.

DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.

DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.

KM Outer membrane; Iron transport; Transport; Tonb box; Signal; Receptor; Complete proteome.

KW Complete proteome.

FT SIGNAL 1 25 POTENTIAL.

FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.

FT SITE 39 44 TONB_BOX.

FT SITE 729 746 TONB_C-TERMINAL_BOX.

SO SEQUENCE 746 AA; 80967 MW; ADD4FB0CB23C3251 CRC64;

Query Match 6.8%; Score 256.5; DB 1; Length 746;

Best Local Similarity 21.3%; Pred. No. 4.9e-10;

Matches 175; Conservative 117; Mismatches 321; Indels 209; Gaps 40;

QY 21 LSSVFAATADLETVHTIGGNSYNAIVTEKNGDYSSFAVYGTGKIPASLREIPGS--VS 78

DB 14 LSSCLIA-----NAVHAAGQ-----GDGSVIELGEGTVAVTAQTEETKQAPGV 56

QY 79 IITNQVKDR-NVDFPDQARKTPGLRVLSD-----DG- 111

DB 57 IITAEIAGRPSPNSDLSQIRTPMGNLGNSSSGRGNNRQIDIRGMPEWTLILVDK 116

QY 112 ----RSV- YA-RGYEYSYINIDGLPAQMSINGTLPNLPAPDRVFMVMPGSLFDSGE 165

DB 117 PVSSRSVYVGRGERDSGDTNWVPADQ-----VERIEIVIGPAARYGNGA 164

QY 166 MGIIVLVKRRPTKAFQGA--AGGCTH-----QKAEADVSG 203

DB 165 AGGVNVITTKQAGETHGMLSVSNFPQHKAGASERMSFGINGPITENLSTRVYGNLIK 224

QY 204 SLNSDSVGRVMAQTVG--ASPRPAEKNN-RHETFYAADWDINDPYL--GAGLYOQ 258

DB 225 TSDSDWDINAGHSNRGTQAGTLPRAGRGVANNKQIDGLSLRRLTEQTLFEFAG-SRQ 283

QY 259 RHLAPYNGLPADANNKLPPLPQHVFVADWNKFKNNSHDVFPADLKHYFGNG-GYKVGMR 317

DB 284 GNIYTGDTQNTSNVYVKQM-----LGHETNMVBETYSVTHRGEMDFSSLAY----LQ 334

QY 318 Y-SDPADSNVAFAGSKLGKTPAGRGCGNTADDACAVGLGETIKQKALADASIRFP 376

DB 335 YEKTRNSRINEGLAGTGEI FDP-NNAGFYTA-----TLRDLTAHGEVNLPL 380

QY 377 RLGNANBEFVIGADVNRFRSTNEQGRRTLYARGGLALNFRSIPQVDLIANARKGVGY 436

DB 381 HIG-YEQTLITLSEWTEQQLDDPSSNTQNTBERG-----SIP-----GIAGKN 422

QY 437 HTVAENIDFEGYKSTFHPADGLSLIGGRLGHYKIES--GECKTLHKASKTKFTGY 493

DB 423 RS-SSSARIFSLPEDIINELMPGTMVLGRLWHDHDI VGDWMSPLNLSHALTERVTIK 481

QY 494 AGA-----VYDLDNNNSLYLSQLYTPQTNL---DADKLLKPRGNGPEVGYKSS 542

DB 482 AGIARAYPKNLYQNPDLVLSRGGCGYSTCYLRGNDG--LKAETSVNKEIGIE-- 537

QY 543 YMDRLNARVSFYRMKDKY-AAAPINP-----NNKTRYA-----ALGKRVMEGVETE 589

DB 538 YSHDGLVAGLTFRNDYKKNKIESGLSPVDHASGKGVANNAIYQWENVPKAVVEGLE-- 595

QY 590 ISGAVT-----PKQIHAGSYLHSQITKASNSRDGIFLLMPKHSANLWTTYQVTPEL 643

DB 596 --GITLPLPADGLKMSNNLT-----MGSKNKETGVLSTPRXTLNSMLDMQWQTDLL 647

QY 644 TTGGGVNAM-----SGITSSAGMAHAGIATFPAMAAIYRTPYKLIQINADNIF 691

DB 648 SIQATVTWYKQKPKKYDYHGDVTVGSANDQLSPYAIAGLGGTYRLSKNLSGAGVDNLF 707

QY 692 NNHYVARVG-----GANTENIPGSERTTANLRYSF 722

DB 708 DKRLF-RAGNAGGVGIGDAGATTNEPG--RFTYTSLTASF 746

RESULT 16

YEGA_HAEIN STANDARD; PRT; 345 AA.

AC 086241; 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypothetical tonb-dependent receptor H11466.1.

GN H11466.1.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Rd / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,


```

RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; U32825; AAC23131.1; -.
DR TIGR; H11466.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Hypothetical protein; Receptor; TonB_box; Complete proteome.
FT SITE 325 345 TONB C-TERMINAL BOX.
SQ SEQUENCE 345 AA; 38744 MW; EAF83C55FCC9CB2A CRC64;

Query Match 5.9%; Score 221.5; DB 1; Length 345;
Best Local Similarity 24.1%; Pred. No. 4.2e-08;
Matches 82; Conservative 58; Mismatches 129; Indels 71; Gaps 16;

QY 435 YGHTVATE-----NLDFEFGYKSTFHPADGLSLGGGRIGHY-----K 473
DB 25 YINPVAIKNNENNAVQFNHLKTLGLYQNTYFTDNFIITGLAYEYDQVVGSRSLKN 84
QY 474 IESG-----EGTLHKASKTGTGAVYDLNDNLSLSQLYTPQTN--LDADGKL 526
DB 85 IRSGVLAQDKGKLLYQ-----LGSVYKFTPNIAFPFNHAEFRPQNNRTLIINGE- 134
QY 527 LKPRGNOFEVGYKGYMDRLNARVSFYRMKDKNAAPLNPNKKTRVYALGKRVMEGV 586
DB 135 LPAEQGSFETGLK--YENAYLNATVALENINKRNVAETVNVNG--TNELQIVGKQSRGI 191
QY 587 ETEISGAVTPKQIHAGYSYLSQIKTASNRD-----DGIFLLMPKHSANLWTTQY 638
DB 192 EFDLNGQLTDNLSIAANYTY--TKVKNLENHNKLVAGKQLSGV---PKHQASLFLAYN 245
QY 639 VTP-----ELTIGGVNAMSGITSSAGMHAGGY-----ATFDMAAYRFT--PKLKLQIN 686
DB 246 IGFEFDGPNIRVGCGARYLGSWYANNYTYKAYLPOAIVYDTFYADTKISGKYSFQLN 305
QY 687 ADNINFRHY-ARVGANTFNIP--GSERTWANLRSYF 722
DB 306 GKNLSNKKVYSPSTSGNASRTLIPVALGYAREVILNTKIEF 345

RESULT 17
FATA VIBAN STANDARD; PRT; 726 AA.
AC P11461; P19830;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric anguibactin receptor precursor (OM2).
GN FATA.

```

```

OS Vibrio anguillarum (Listonella anguillarum).
OG Plasmid pJMI.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.
OX NCBI_TaxID=55601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=775;
RX MEDLINE=88139336; PubMed=2830268;
RA Actis L.A., Tolmasky M.E., Farrell D.H., Crosa J.H.;
RT "Genetic and molecular characterization of essential components of
RT the Vibrio anguillarum plasmid-mediated iron-transport system.";
RL J. Biol. Chem. 263:2853-2860(1988).
RN [2]
RP SEQUENCE OF 637-726 FROM N.A.
RC STRAIN=775;
RX MEDLINE=90185247; PubMed=2311935;
RA Farrell D.H., Mikesell P., Actis L.A., Crosa J.H.;
RT "A regulatory gene, angR, of the iron uptake system of Vibrio
RT anguillarum: similarity with phage P22 cro and regulation by iron.";
RL Gene 86:45-51(1990).
CC -!- FUNCTION: ANGUIBACTIN RECEPTOR, ESSENTIAL COMPONENT OF THE IRON
CC UPTAKE SYSTEM OF V. ANGUILLARUM.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; J03529; AAA91581.1; -.
DR EMBL; M34504; AAA79859.1; -.
DR PIR; B29928; B29928.
DR PIR; PQ0051; PQ0051.
DR InterPro; IPR000531; TonB_boxC; 1.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;
KW Transport; TonB_box; Plasmid.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 726 FERRIC ANGUIBACTIN RECEPTOR.
FT SITE 49 56 TONB BOX.
FT SITE 709 726 TONB C-TERMINAL BOX.
SQ SEQUENCE 726 AA; 78894 MW; 88B67A219395B154 CRC64;

Query Match 5.7%; Score 215.5; DB 1; Length 726;
Best Local Similarity 21.6%; Pred. No. 3.1e-07;
Matches 163; Conservative 115; Mismatches 338; Indels 137; Gaps 37;

QY 9 RINMTAATVLAALS---SSVFAAQATADLETVHIKQSYNAIVTEKNGDYSSFAVTVGCK 65
DB 22 RIALVALQISASALPISITHAEQAD-ESITVYQGAN-EAYAAGKISKASSIGH-LGDK 78
QY 66 IPASLREIPQSVSITITQQVKDRNV-DTFDQARKTPELRVLSNDDG--RSSVYARG-- 120
DB 79 ---DFLDTFPNAIGYTDKHQIQDHAQDISVISADPSV-FTSGETGLNKESPKIRGSS 134
QY 121 EYSEYNIDGL-----PAQMQSINGTLNLPAPDRVENVRGPSGLFDS---SGEMGGIVNLV 173
DB 135 DIGDVMFNGLYGIAPIYRSS-----PEM--YQRIDVLKGPASLLNGMPNGSVGGSINLV 187
QY 174 KRPTKA-----FQGHAAAG--FGTHKQYKAED---VSGSLNSDGSVGR 214
DB 188 TKRAQEAIPITFTGYMSDSQFGGHIDIGRRFGENEGVGFNFRDGDASVDGQSKA 247
QY 215 VMAQTGVGASPRPAEKNRHHETFYAAADWDINPDTVLGAGYLYQQRHLAPYNGLPADANNK 274
DB 248 QLAS-----LSLDRNDIALIEADLYFSTERVDPGNRGLSTASGVD 288

```



```

QY 275 LPSLP-QHYFVGADW--NFKFKNSHVDFADLKYFGNGCYGKCYKRYSDRDASTAFAG 331
DB 289 VSPSPSDTLTSSPMAYNSEDGMMIRAEILDSNSVTAYGAVGA--SRTDFSPVPORY 346
QY 332 SKTGKMTAPGRPCNTADKACVIGT--BIKKALAFDASYSRPRLTANEPIYIGAD 390
DB 347 KIID-----DSGTLEVSLSGVALBSKRTSGEIGIRSPDTGTIEHYLVANST 393
QY 391 YNFRSTNEGRTTLTYARAGGLALNEFRSI--POVDLIANAKRGVSGHATENIDFEG 448
DB 394 Y--FRDKNDSPTGNPNPGMNINYPWGPEDSTYDN-----YELPVDSTQISFG 444
QY 449 IYKSTFHPADGSLIGGG-----RIGHYKIESGEGKTLH-----KASKTFYVAGA 496
DB 445 V-----ADTLISANGKQVTLGLRHOSIDYEGVTWNGNAFPPTKLSESTYPAIVA 496
QY 497 VYDLNNNSLYLSLSQLYTPQTLDD-----GKLLKPRQNGPEVGYKSGYMDRLNARY 552
DB 497 LYKVSQSVSLYGYTGLTNGKTAGSGAANVGAFEPQKTKQTEAGLKLD--MNDPAHT--F 554
QY 553 SFYRMKDKNAAAPLNPNNKTRRYAALGKRVMEGVEITEISGAVTPKQIHAGSYLSHQIK 612
DB 555 SLEIFKPKPGYQDPDNT-----IYSPGEGQRNRIEMGFYCTVLEDYTLNGIAYTDAEIT 610
QY 613 TASNSHDDG--IFLLMPKHSANL--WTTYQVTELTIGGCVNAMSQ--ITSSAGMHAGY 666
DB 611 KATDVTEGKQATKLPDLQKALAEW--NLPMVQLTLIGQANMSQYIDAQNTQISLAQ 669
QY 667 ATPDAAAYRFT---FKLQIQINADNIFNHHY 696
DB 670 TIFDLGARYNSTIANOSVIMRLVANNVTDAAYW 702

RESULT 18
IRGA_VIRCH STANDARD; PRT; 652 AA.
ID IRGA_VIRCH STANDARD; PRT; 652 AA.
AC P27772; Q9KUP0;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron-regulated outer membrane virulence protein precursor.
GN IRGA OR VC0475.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OS Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_Taxid=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=93023868; PubMed=1406279;
RA Goldberg M.B., Boyko S.A., Butterton J.R., Stroebner J.A.,
RA Payne S.M., Calderwood S.B.;
RT "Characterization of a Vibrio cholerae virulence factor homologous to
the family of TonB-dependent proteins.";
RL Mol. Microbiol. 6:2407-2418(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gaim M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Basse S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Frazer C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
RN [3]
RP SEQUENCE OF 1-151 FROM N.A.
RC STRAIN=classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=91072235; PubMed=2174861;
RA Goldberg M.B., Boyko S.A., Calderwood S.B.;

```

```

RT "Transcriptional regulation by iron of a Vibrio cholerae virulence
gene and homology of the gene to the Escherichia coli fur system.";
RL J. Bacteriol. 172:6863-6870(1990).
CC -1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING
CC FERRIC VIBRIOBACTIN. AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC VIBRIOBACTIN TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY
CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, U72152; AAC4766.1; -.
CC EMBL, AE004134; AAF93648.1; -.
CC PIR, A37834; A37834.
CC PIR, S25265; S25265.
CC HSRP, P03825; IFEF.
CC TIGR, VC0475; -.
CC InterPro: IPR000531; TonB boxC.
CC Pfam: PFO0593; TonB boxC.1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE: PS0156; TONB_DEPENDENT_REC_2; 1.
CC Virulence; Outer membrane; Iron transport; Transport; TonB box;
CC Signal; Receptor; Complete proteome.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE
CC PROTEIN.
CC SITE 33 40 TONB BOX.
CC SITE 635 652 TONB C-TERMINAL BOX.
CC CONFLICT 294 294 D -> G (IN REF. 1).
CC CONFLICT 448 448 K -> Q (IN REF. 1).
CC CONFLICT 502 502 A -> T (IN REF. 1).
CC SEQUENCE 652 AA, 71669 MW, A95F82EFC072EC93 CRC64;

Query Match 5.7%; Score 214; DB 1; Length 652;
Best Local Similarity 19.6%; Pred. No. 3.4e-07;
Matches 146; Conservative 98; Mismatches 305; Indels 196; Gaps 31;

QY 57 SPAYVTGKIPAS-----LREIPQSVGIIITNOQYKDRVDF 93
DB 10 SLSTVLGLMFSASAFQADATKTDETWVTAAGYAQVIONAPASISVISREDELSRYRDV 69
QY 94 DQARKTPLGLVLSNDDGRS--SVYARGVEYSEYINIGLPAQMOI-----NGTLP 142
DB 70 TDALKVPEVYTVGGDDTTDISIRGMSNYTLIVDGKQTSRQRPNSDPRIEGWL 129
QY 143 NLEAFDVEYMRGSGLFDSGEMGIVNLVRRPTKAFQGHAAAGFTHQYKABADVS 202
DB 130 PLQAIIRIEVIRGPMSTLYGSDAIGVINITRKQDOQSGNVQSTVVOENRASDEQS 189
QY 203 GSINSGSRGRMAQTVASRPAKNNR-----ETFAAADMDNDNP--TVLGAG 253
DB 190 ANEFVTGPLSDALSLQVYQTTQRDEDEIEHGYSKLSLTSKLVYQINPHQQLLEAG 249
QY 254 YLYQQR-----HLAPNGLPADANNKLPSPQHFVFG--ADWNKFKMNSHDVPADLK 303
DB 250 VSAQDEBNVNGSAGSSGCGRTCSNTDQYRR--NHVAVSHQGDW-----QDV----- 295
QY 304 HFFGNGYKGVKMGYSRDRADS-----NYAFAGSKLGMKTPAGRGPCNTADKACAVGL 357
DB 296 -----GQSDTYLYQYBENTNKSREMSIDNTVF--KSTLVAPIGE-----MLSF 336
QY 358 GTEIKKALAFDASYSRPRLTANEPIYIGADYNFRSTNQGRTTLTYARGLALNEFR 417
DB 337 GVEGKHSLE-----DKTSNKI-----SSRTHLSNTQW-----AGFIEDWA 373

```



```
QY 418 SIPOVDLIANARKGVGYSHYVATENLDFEGLYKSTFHP-----ADGLSLGGRLG 470
Db 374 LAEOPRLTFGGR-----LDHDXNYG-SHFSRPRYGVWNLDPLTWVRGG--- 415
QY 471 HYKIBSG-EGKTLHKASKT--KFTGVAGAVY--DLNDNNSLYLSLSOLYTPQTNLDADG 524
Db 416 ---VSTGFRAPOQREVTPWQVSG--GGNIYGNPDLKPETSKINKELSLMYSGSGLAAS- 470
QY 525 KLLKPRQGNQFEVGYGSGYMDRLNARVYFRMKDKNAAPLN-----PNN---KKTRYA 576
Db 471 -----LTAFHNDKDKITRVACPANICTAGBNQWGAAPTYYV 507
QY 577 ALGKRVMEGVETEISGAVTPKQIHAGISYLSHQIKTASNSRDDGIFLLMPKHSANLWTT 636
Db 508 NIDEAEYGAETLSLPITESVELSSSYTYTHSEQKSGNFAGR--LLQLPKHFLNANLS 565
QY 637 YQVTPPELTIGGVN-----ANSGITSSAGMHAGGYATFDAMAAYRFTPKLKLQINADNI 690
Db 566 WQTDRLNSWANLVNRYKEMQEPGGASNDDFIAPSYTFIDTGVYALTDTATIKAAVNL 625
QY 691 FNRHY-YARVGGANTFNIPGSRTW 714
Db 626 PQEYNYAEYG-----YVEDGRYY 645

RESULT 19
YNCN ECOLI STANDARD; PRT; 700 AA.
AC P76115;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE Probable tonB-dependent receptor yncD precursor.
GN YNCN OR B1451.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474 (1997).
CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000242; AAC74533.1; -.
CC EcoGene; EG13774; yncD.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;
KW Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 700 PROBABLE TONB-DEPENDENT RECEPTOR YNCN.
FT SITE 680 700 TONB C-TERMINAL BOX.
FT SEQUENCE 700 AA; 77260 MW; 7B3B96C6ABA48FE4 CRC64;
SQ
```

```
Query Match 5.7%; Score 214; DB 1; Length 700;
Best Local Similarity 20.0%; Pred. No. 3.8e-07;
Matches 156; Conservative 114; Mismatches 317; Indels 192; Gaps 31;

QY 5 MSVFRINNTAATVLAALSSSVFAAQTADLETVHIKGRSNAIVTEKNGDYSSFAVTGT 64
Db 1 MKIFSVRQTVLPALLVLSPVVFAA-----DEQTVIVSAAP 35
QY 65 KIPASLREIPOSVSIITNQVK--DRNVDTFDQLARKTPGLRVLSNDD----- 110
Db 36 QWSEL-DTPAAVSVVDGEEMRLATPRINLSESLT-GVPGLOVQNRQNYAQDLQLSIRGP 93
QY 111 GRSSVYARGYSEYNIDGLPAQMSINGTLPN--LFAFDRVEVMRGP-SGLFSSSGEMG 167
Db 94 GSRSTY--GIRGIRLYVDGIPATMPDGGQTSNLDLSVQNVLEVLRGPFSAlyGNAS--G 149
QY 168 GIVNLVRKPTPKAFQGAAGFGTHKQYKAEADYSGSLNSDGSVGRVMAQTVGASPRPA 227
Db 150 GVMNVTTQTGQOPTIEASSYSGFSGWRVYGLKATGA-TGDTQTPGDV----- 196
QY 228 EKNRHETFYAAADWDINPDTVLGAGLYQ---QRHLAPYN-GLPADANNKPLSPQHVF 283
Db 197 -----DYTVSTTRFTTHGYRDHSGAQKLANAKLGVRIDEASKLSLI----- 238
QY 284 VGADWKEFMNSHDVPADLKHFGNGGYGKVMRYSDRSDANSYAFAGSKLGMKTPAG-- 341
Db 239 -----FNSVDIKAD-----DPGGLTKAEWKANPQAPRAEQYDTRKTIKQOAGLR 284
QY 342 -RPGCNTADDKACAVGLG-TEIKOKALAFDASYSRPFRLGN--TANEPVIGADYNRRPST 397
Db 285 YERSLSRDDMSVMYAGERETTYQSIHPAQLNPSHAGGVITLQRHYQID-SRWTHR 343
QY 398 NEOGRTTLYARGGLALNEFRSIPQVDLIANARKGVGYSHYVATENLDFEGLYG---- 451
Db 344 GELGVPTFTTGT---LN-----YENMSENKRGYNPNRNLNSGMPYEQKGLERDE 390
QY 452 -----KSTFHPADGLSLGGRLG-----HYKIESGEGKTLHKASKTKPTG 492
Db 391 RNLMMNIDPYLQTMQQLSEKLSLDAGVRYSSVWFDSNDHY-VTFNGDSDSDGASVHKWLP 449
QY 493 YAGAVYDLNDNNSLYLSLSOLY-TPQTN---LDADGKL-----LKPRQGNQFEVGYKSGY 543
Db 450 AGSKYAMTDANNIYLAAGRGFTPTNELSYRADGSGMNLGLKPNSTNDTIEIGSKTRI 509
QY 544 MDDLRLNARVSPYRMKDKNAAPLNPNKTKTRYAALGKRVMEGVETEISGAVTPKQIHAG 603
Db 510 GDGLLS--LALFQTDTDDEIVVDSSSGRRTTYKNAKTRRQGAELAWDQRFAGDFRYNAS 567
QY 604 YSYLHSQIKTASNSRDDGIFLLMPKHSANLWTTVQVTPBELTIGGVNAMSITSSAGMHA 663
Db 568 WTWLDATYRSNVCNEQDCNGRMFPIARNMGF-----ASIGYVPEDGMYA 612
QY 664 GGYATF--DAMAAYRFTPK-----LKLQINADNIENRHYARV 699
Db 613 GTEARYMGDINADDENTAKAPSYTLVGLFTGYKYNHYHNLTVDLFRVDNLFDEKYVGSV 671

RESULT 20
BTUB_SALTY STANDARD; PRT; 614 AA.
AC P37409;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vitamin B12 receptor precursor.
GN BTUB OR STM4130.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L72;
```



```

RX MEDLINE-93079446; PubMed1448622;
RA Wei B.Y., Bradbeer C., Kadner R.J.;
RT Conserved structural and regulatory regions in the Salmonella
RT Typhimurium brub gene for the outer membrane vitamin B12 transport
RT protein."
RL Res. Microbiol. 143:459-466(1992).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGCSC142 / ATCC 700720;
RX MEDLINE-21534948; PubMed1167609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
CC -!- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC -----
CC DR EMBL; M89481; AAA27031.1; -
CC DR EMBL; AB008893; AAL22968.1; -
CC DR StyGene; SG10030; brub.
CC DR InterPro; IPR000531; TonB boxC.
CC DR Pfam; PF00593; TonB boxC_1.
CC DR PROSITE; PS00430; TONB DEPENDENT REC_1; 1.
CC DR PROSITE; PS01156; TONB DEPENDENT REC_2; 1.
CC DR Outer membrane; Cobalt transport; Transport; TonB box; Signal;
CC KW Receptor; Complete proteome.
CC FT SIGNAL 1 20 BY SIMILARITY.
CC FT CHAIN 1 614 VITAMIN B12 RECEPTOR.
CC FT SITE 26 33 TONB BOX.
CC FT SITE 597 614 TONB C-TERMINAL BOX.
CC FT CONFLICT 81 81 G -> R (IN REF. 1).
CC FT CONFLICT 134 134 R -> P (IN REF. 1).
CC FT CONFLICT 183 183 N -> I (IN REF. 1).
CC FT CONFLICT 280 280 R -> S (IN REF. 1).
CC SQ SEQUENCE 614 AA; 68525 MW; 9F51F601A615FD62 CRC64;

```

```

Query Match 5.4%; Score 205; DB 1; Length 614;
Best Local Similarity 21.1%; Pred. No. 1.3e-06; Indels 206; Gaps 39;
Matches 161; Conservative 115; Mismatches 280;

OY 15 ATVTALSSVFAQTADLETVAHIKQORSYNAIVTEKNGDYSFAVTGKIPASIREIP 74
DB 5 ATTLTAFSVTAFSAMQD-----TSPDTLV-----VIANFQGP 38

OY 75 QS-----VSIITNQVYKDRNVDTFDQARKTPGLRLVLSNDG--RSSVARGYESE--Y 125
DB 39 RSVAVLAPVITVTRQDIERMSTSVNDVLRPLPGVDIAQSGACQNSIFRTGNSSHVLV 98

OY 126 NIDGLPAMQMSINGTLPLNFAF-----DREVVVRGSPGLFDSGCGGTYNLVRRKPTKA 180
DB 99 LIDGVNLINAGVSGS--ADLSQFPVSLVQRLEYIKRGRSAIYGSDAIGGVNITTTDNEG 157

OY 181 FQGHAAAGFTKQRYKAEDVDSGLNSDSVGR-----VMAQTVGASPRPAEKNN 231
DB 158 TELTAGMGSNSYQNDISTQOQUGENTRATLIGDYEYTKGFDVAVAGGTGMQADPRDGR 217

OY 232 RHETFAAAD-----WDINPDVYLGAQYLYQORHLAPIN-GLPADANNTLPLPGHFVY 284
DB 218 LSKTLVGALEHTFSRWS--GFVRGVDYDNRIDYDAYVSPGSPLLIDTRKLVG--OSWDA 272

OY 285 GADWNKFNKNSHDV--FADLKHYFNGGYGVKVMRYSDRSDASNYAFAGSKLGMKTPAGR 342
DB 273 GHFNGERIQLQSLVSSYSHSKDYNDPHYG-----RX-DTSA----- 308

OY 343 PGCNTADDKACAVGLGTEIKQKALAFDASYSRPRFLGNTANEFGADVNRERST----- 397
DB 309 ----TLD-----EMKO-----YNVQWTSVVVGH--GNVGAGVDWQKOTTPGTGY 348

OY 398 ----NEQGRITLYARGGLALNFRSIPQVDLJANARKVGRVSHVATENLDFGIYGS 453
DB 349 VPEGVDQKNTGYVLTGLQGLDP-----TLEMAAR-----SDNSQFGRHG-- 389

OY 454 TPHPADGLSLGGGRIGHYKISGEGKTLHKASKTFGTGAGAVYDLNDNNSLYSLSQL 513
DB 390 TWQTSAGWPFIS-----YRPLASYG-TSYKAP-----NLGGL 421

OY 514 YTPQTNLADGKLLKPRQNGPEVGYKSYMDRLNARVSFPRMDKNAAPLNPNKKT 573
DB 422 YGYGPN-----LNPEKSKQWGAFFEG--LTAGVSWRISGYR-NDINMIDYDHLQK- 472

OY 574 RYAAAGKRWMEGVETEI---SGAVTPKMQIH--AGVSYLHSQIKTASNSRDPGIFLMPKH 629
DB 473 -YNEBKARIKGIEATANFTGPPLT-----HTVSYDYV-----DARNAITD--TPLEPR 518

OY 630 SANL-----MTTYQVTPELT---IGGVNAMSGLITSSAGMHAGYATPDMAAAYRFTPK 680
DB 519 SKQMAKYQLDMDVDPDWMGTQYILGRSDSDYSANPRTVAKMGVSLMDLYAVYVTSH 578

OY 681 LKQLINADNIPNRHYYARVGANTENIPGSERTWTANLRYSF 722
DB 579 LTVRGKIANLPKDYETVYG-----YOTAGREYTLGSS--YTF 614

RESULT 21
FECA_ECOLI STANDARD; PRT; 774 AA.
ID AC P13036;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Iron(III) diclitate transport protein fecA precursor.
GN FECA OR B4291.
OS Escherichia coli.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 34-40.
RC STRAIN=B;
RX MEDLINE=8827855; PubMed=2836368;
RA Presler U., Staudenmaier H., Zimmermann L., Braun V.;
RT "Genetics of the iron diclitate transport system of Escherichia
RT coli."
RL J. Bacteriol. 170:2716-2724(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE OF 1-428 FROM N.A.
RX MEDLINE=91072220; PubMed=2254251;
RA van Hove B., Staudenmaier H., Braun V.;
RT "Novel two-component transmembrane transcritpion control: regulation
RT of iron diclitate transport in Escherichia coli K-12."
RL J. Bacteriol. 172:6749-6758(1990).
RN [4]
RP SEQUENCE OF 731-774 FROM N.A.

```


RC STRAIN=K12;
RX MEDLINE=89213950; PubMed=2651410;
RA Staudenmaier H., van Hove B., Varaghi Z., Braun V.;
RT "Nucleotide sequences of the fecBCE genes and locations of the
RT proteins suggest a periplasmic-binding-protein-dependent transport
RT mechanism for iron(III) dicitrate in *Escherichia coli*.";
RL J. Bacteriol. 171:2626-2633(1989).
CC -!- FUNCTION: FECA is the OUTER MEMBRANE RECEPTOR PROTEIN IN THE
CC IRON(III) DICITRATE TRANSPORT SYSTEM.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC AND INDUCTION BY THE IRON LEVEL MEDIATED BY THE FUR PROTEIN
CC DICITRATE COMPLEX MUST ENTER THE PERIPLASM WHERE IT BINDS TO A
CC TRANSMEMBRANE PROTEIN, WHICH REGULATES DIRECTLY OR VIA A FURTHER
CC INDUCTOR, TRANSCRIPTION OF THE FEC GENES.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20981; AAA23760.1; -;
DR EMBL; U14003; AAA97187.1; -;
DR EMBL; AE000499; AAC7247.1; -;
DR EMBL; M63115; AAA23768.1; -;
DR EMBL; M26397; AAA23761.1; -;
DR PIR; JY0022; QRECEA.
DR EcoGene; EG10286; fecA.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 33
FT CHAIN 34 774
FT SITE 56 63
FT SITE 757 774
FT TONB C-TERMINAL BOX.
FT CONFLICT 16 16
FT T -> A (IN REF. 1).
FT CONFLICT 190 190
FT A -> T (IN REF. 1).
FT CONFLICT 300 300
FT T -> M (IN REF. 1 AND 3).
FT CONFLICT 357 357
FT W -> R (IN REF. 1).
FT CONFLICT 444 444
FT L -> V (IN REF. 1).
FT CONFLICT 749 749
FT R -> A (IN REF. 1).
SQ SEQUENCE 774 AA; 85321 MW; 16B5B510276C3B09 CRC64;

Query Match 5.4%; Score 205; DB 1; Length 774;
Best Local Similarity 20.1%; Pred. No. 1.8e-06;
Matches 170; Conservative 108; Mismatches 316; Indels 250; Gaps 39;

QY 18 LAALSSVFAAQ-----TADLETVHIKQORSYNAIVTEKNGDY- 55
DB 24 LAGLSFAFAAQNTPAGSLDKALNQYAAHSGFTLSVDASLTRGQS-NGL-----HGDYD 78
QY 56 -----SSPAVTV-----GKTPASLRREIPQSVS 78
DB 79 VESGLQLDGGSLQVKPLGNNSWTLEPAPAPKEDALTIVGDWLDGARENDVFEHAGARD 138
QY 79 IITNQVKDRNDVTFDQARKTGLRLVSLNDGDRSSVARGVESEYN-----ID 128
DB 139 VIRREDFAKTGATWKEVLNRIIPGVSAPENNGTGGSHDLAWNFGIRGLNPRLASRSTVLMD 198
QY 129 GLPA-----QMSINGTLNLFADFVEVNRGPGSLFDSSGEMGGVNLVVRKPTKA 180
DB 199 GIPVPFAPYQQLSLAPVSLGNM---DAIDVVRGGAVRYGQSVGVGVNFV---TRA 251
QY 181 FQGHAAAGFTGHQYKAEADVSGSLNSDGSVGRVVAQTVGASPRPAEKQNRHET----- 235

Db 252 IP-----QDFGIEAGVEGQLSPSTSS-----QNNPKETHNLMV 283
QY 236 -----FYA---AADWDIN-----PDTVLGAGLYQQRH-----LAPYNG---L 267
DB 284 GGTADNGFGTALLVSGTRGSDMREHSATRIDDLMLKSKYAPDEVHTFNSLLOYDGEADM 343
QY 268 PA-----DANNKLPSPQHVFGADWNKFKMNS--HDVFADLKHVFGNGGYGKGVNRY 318
DB 344 PGLSLRADYDADRQWQSTRPYDRF-----WGRRLKASLSGLGQFQDSQHKFNIGQFYQTILR- 398
QY 319 SDRDADSNYAFAGSKLGMKTPAGPGCGCTADDKACAVGLGTEIKQKALAPDASRPRRL 378
DB 399 -----SGYLEQGRKITL-----SPRNYWVRGIEPRYSQIPMI 430
QY 379 GNTANEFVIGADYNFRSTNEQGRITLYARGGLALNEFRSIPQVDLIANARKVGRGYSHT 438
DB 431 GPSAHEVGVGRYLN-ESTHEMRYTATSSQLPSG---SSP---YDRDTSGTEAHAW- 482
QY 439 VATENLDEFGYIGKSTPHPADGLSLIGGRLGHVYKISGEK-----TLHKASKTKFTGYA 494
DB 483 ----YLDCKIDIGNWTITP-----CMRPEH--IESYQNNALITGTHEEVSYNAPLAL 528
QY 495 GAVYDLNDNNSLYL-----SLSQLYTPQTNLDADCKLLKPRQGNQFVYKSGSYMDRLNA 550
DB 529 NVLYHLTDSMNLVYANTGSGFTGVQYSLGKAVQSGNVEPEKARTWELGTR--YDDGALTA 586
QY 551 RVSPYRMKDKNAAAPLNPNNKKTRYAALGKRVMEGVETEIS---GAVTP---KWQIHAGY 604
DB 587 EMGLFLINPNN---QYDSNQNTNDVTARGKTRHTGLETOARYDGLTPTLTDNVSIIYASY 643
QY 605 SYLHSQIKTASNRDDGIFLLMPKHSANL-----WTTYQVTPTELTIGGVNAMSQIT 656
DB 644 AYVNAEIREKGDITYGN-LVPPSPKHKGLTGVYKPGNW-TFNLSDFGSSQFADNANTVK 701
QY 657 SSAGMHAG--GYATFDMAAYRTPK---LKLQINADNIENRHYRYARVGANTFNI-PG 709
DB 702 ESADGSGTRIFGFMWGARVAIDFGPQWADNLAFGVKNIFDQDYFIRSYDDNNKKGIYAG 761
QY 710 SERT 713
DB 762 QPRT 765

RESULT 22
HEMR YEREN
ID_HEMR_YEREN STANDARD; PRT; 687 AA.
AC P31499; 1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Hemrin receptor precursor.
GN HEMR.
OS Versinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Versinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RX MEDLINE=93049186; PubMed=1425573;
RA Stojiljkovic I., Hantke K.;
RT "Hemin uptake system of *Versinia enterocolitica*: similarities with
RT other TonB-dependent systems in Gram-negative bacteria.";
RL EMBO J. 11:4359-4367(1992).
RN [2]
RP REVISIONS.
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;
RA Stojiljkovic I.;
RL submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.


```

CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X68147; CAA48250.1; -
DR PIR, S28042; S28042.
DR InterPro, IPR000531; TonB_boxC.
DR Pfam, PF00593; TonB_boxC; 1.
DR PROSITE, PS00430; TONB DEPENDENT REC 1; 1.
DR PROSITE, PS01156; TONB DEPENDENT REC 2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal;
KW Receptor.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 687 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 670 687 TONB C-TERMINAL BOX.
SQ SEQUENCE 687 AA; 75226 MW; 69092EF97F5DC08A CRC64;

Query Match 5.3%; Score 200; DB 1; Length 687;
Best Local Similarity 20.3%; Pred. No. 3.3e-06;
Matches 158; Conservative 96; Mismatches 322; Indels 204; Gaps 32;

QY 8 FRINMTAATVLAALSSVFAAQADAEYTHIKQGRSYNAIVTEKNGDYSSFAVTVGTKIP 67
DB 9 FPMSPSLIACTSLAVQAADATSSDTQ-NSKKRIADTVATGERSFSF----- 58
QY 68 ASLREIPQSVSIITNQVDRNVDTPDLARKTPGLRVLSNDGR--SSVARGEYSE 124
DB 59 ----EAPMAVTVVEATPTPTSETATSDMLRNIPGLITVGS--GRVNGQDVTLRIGTKQG 112
QY 125 Y--NIDGLPAQMS--INGTLFNLFAFDRVENRGPSPGLFDSGEGVY-----NL 172
DB 113 VTLVVGIRGQDTGHLNSTFLDPAVKVEIVRGPALLYSGGALGVVSYETVDAADL 172
QY 173 VKRPRKPAQGHAAAGFGTH-----KQYKAADVSGSLN---SGSVRGRVMAQTGA 222
DB 173 LPPGQNSGVRYSAAATGHSFGLGASARFRTDVGIIISFGTRDIGNIR-----OSDGF 227
QY 223 SPRPAKNNRHETFYAADMDINPDTVLGAGVLYQGRHLAPYNGLPADANNKLPSPQHV 282
DB 228 N--APNDETISNVLAKGTRRIDQIOSLSANLRY-----YNSALEPRKNPOTSAASST 277
QY 283 FVGADNNKFRKNSHDVFAADLKHYFGNGYGVKVMRYSDRDADSNVAFAGSKLGKMK----- 337
DB 278 NMTDSTLIQRDAQ-----LKNIKRPLDDEWMLNATQVYVSEIN 318
QY 338 -TPAGPFGCNTADDKACAVGLTEIKOKALAPDASYSRPRLCNT-----ANER 385
DB 319 ABBPGQTP-----EERGKQTTKGKGLNRTTLFTDSFASHLLTGYTAAYKQEQTPSGATES 373
QY 386 VLGADVNRFRSTNEQRTTLVYAGGLALNEFRSIPQVDLIANNR-KGVGYSTIVATENL 444
DB 374 FPOADL-KRFGSGLQDEITL-----RDLP-VSIIAGTRYNNRSSSEGVADVDA 420
QY 445 DEFGIYKSTFHPADGLSLIGGRLGHYKLESGEKTLHKASXTKFTGYA-----GA 496
DB 421 DKSSSRGAVSVPTDMLMFG-----STVAQAFRAPTMGE 454
QY 497 VYDLADNNLSLYLSQLYPQTNLADGKLKPRQGNQFEVYKGSYMD-----DRINAR 551
DB 455 MY-----NDSKHFSMNIMGTLLTYWVPNPLKBTETNETQEGYGLFNFNDLMAEDDLQFK 510
QY 552 VSPFYRMKDKNAAPLPLPNKKTRFYALGKRVMEG-----VEVEISGATTPPK 558
DB 511 ASIF--DTNA-----KDYISTGVTMDFGFGPGGLYCKNCSTYSTINIDRAKIWM 557

```

```

QY 599 QIHAGSYTHSQIKTASNRDQIFLLMPKHSANLM-----TTYQVPELITGGGVNM 652
DB 558 DATMTYQOTQWFMFLGLAYN-KTRG-----XQONTNEWLDTPINDTVTSTDDVPANGFAV 611
QY 653 SGI-----TSSAGMHAGVATFDMAAYRFPFKL---LQINADNIPNNHYARFG 700
DB 612 GWIGTFADRSRVSSSGTPQAGVGNDFVSYKGOEKFQKMTTTVVLAHAFDGYIGPGQ 671

RESULT 23
ID TUTA_ECOLI STANDARD; PRT; 732 AA.
AC P14542;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ferric aerobactin receptor precursor (Cloacin receptor).
GN TUTA.
OS Escherichia coli.
OG Plasmid IncFI ColV3-K30.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F344;
RA Krome W.J.A., Stegehuis F., Koningsstein G., van Doorn C.,
RA Rosendaal B., de Graaf F.K., Oudega B.;
RT "Characterization of the pColV-K30 encoded cloacin Df13/aerobactin
RT outer membrane receptor protein of Escherichia coli; isolation and
RT purification of the protein and analysis of its nucleotide sequence
RT and primary structure."
RL FEMS Microbiol. Lett. 26:153-161(1985).
RN [2]
RP REVISIONS.
RA Oudega B.;
RA Submitted (FEB-1992) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X05874; CAA29297.1; -.
DR EMBL, X05874; CAA29298.1; ALT_INIT.
DR PIR, S01042; S01042.
DR InterPro, IPR000531; TonB_boxC.
DR Pfam, PF00593; TonB_boxC; 1.
DR PROSITE, PS00430; TONB DEPENDENT REC 1; 1.
DR PROSITE, PS01156; TONB DEPENDENT REC 2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid;
KW Receptor.
FT SIGNAL 1 25
FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.
FT SITE 31 38 TONB BOX.
FT SITE 715 732 TONB C-TERMINAL BOX.
SQ SEQUENCE 732 AA; 81014 MW; 0C23879C0B827AE2B CRC64;

Query Match 5.2%; Score 198; DB 1; Length 732;
Best Local Similarity 19.5%; Pred. No. 5e-06;
Matches 174; Conservative 116; Mismatches 256; Indels 346; Gaps 48;

QY 6 FRINMTAATVLAALSSVFAAQADAEYTHIKQGRSYNAIVTEKNGDYSSFAVTVGTR 65
DB 8 TLMALNPILLITMMAA-----VAQQTDDTFVVASANRSRTVA----- 45
QY 66 IPASIREIPQSVSIITNQVDRNVDTPDLARKTPGLRVLSNDGRSSVARGEY 122

```



```

Db 46 -----EMAQTWVIEAELQIOGGKELKDALAQILPGLDVSRSTNYGMNVRGRPL 99
Qy 123 SEYNDIGL-----PAQMOSINGTLPNLFADFEVNRGSPGGLFDSGEMGGIVNLVR 174
Db 100 VVL-VDGVRLNSRSDSRQSDIDP-----FNWHIEVIFGATSLY-GGGSTGGLINIVT 152
Qy 175 KR-----PTKAFQOGHAAAGTGTHK-----
Db 153 KKGQPTWMEFPAAGTKSGPSSSKDHERTAGAVSGGNEHISGRLSVAYQKFGGFDGNGD 212
Qy 194 -----QYKAEADV--SGSLNSDGSVGRVMAQTVGASPRPAEKNEHETFYAAA 240
Db 213 ATLLDNTQTGLQYSDRLDIMGTTINDESRLQITQ-----YKYSQ 255
Qy 241 ---DWDINPDTVLGAGLYQOORHLAPY--NGLPADANNKLPSPQHVFGADVWKNFKMNS 295
Db 256 GDDVGLN-----LGKGFSAIRGTSTFPVSNGLNSD---RIFGTDGHL----- 295
Qy 296 HDVFADLKHVFGNGGKVGKMYRSDRSDSNYAFAGSKLGMKTPAGR----- 342
Db 296 -----ISLQYSDS-----AFLQDEL-----VGQVYVRDESRLRFYVF 326
Qy 343 PCNTADDKACAVGLGTEIKQALAFDAS-----YSRPFRLCNTANE-----FVIGADY 391
Db 327 PTVNA-----NKQVTFSSSQDPTDQYGMKLTLSNKPMDGQWTWGLDADH 372
Qy 392 NFRSTNEQ--GRITLYARGGL-----ALNEFRSIPQVDLLANARKG-----VRGY 435
Db 373 ERFTS--NQMFPLAQASASGGLNKKIYTTGRYPSYDITNLAFLQSGVDINNLTGNG 431
Qy 436 SHTVATEN--LDBFGIY-----GKSTPHPADGLSLGGRLGHYKIESG----- 477
Db 432 VRYQYTENKIDDFIGVAAQOQIGAGKAT--SADAFWRL--SRLRHLFNAGLLMHITEPQ 487
Qy 478 -----EKLTHKASKTKFTGYAGAVDLDNNNSLYLSLSLYTPQTNLDADGKLLKPR 530
Db 488 QAWLNFSGQLELDFPKY-----YGRGIYGAAYNGHLPL-----TKSVNV--SDSKL----- 532
Qy 531 QG---NQREVGVKSGMDDRLNARV--SFYRMDKNAAPLNPNKKTRVYALGKRVMEGV 586
Db 533 EGKVDVSEYELGWR--FTGNLRTQIAAYISIDKSVVA-----NKDLTTSVDDKRIYGV 586
Qy 587 ETEISGAV-----TPKWOIHAGYSYLHSQ-----IKTAGNSRDDGIFLLMPKHS 630
Db 587 E---GAVDYLIPDWDSTGVNFVNLKTESKVGNTWQKVDVKTASPSKA----- 631
Qy 631 ANLWTTYQVTPBELTTCGGVNAMS-----GITSSAGMHAGGYATPDMAAYRFTPKL 681
Db 632 ----TAY-----IGWAPDPWSLRVQSTTSFVDSDAQYKVDGTYTVDLLGSYQL--PVG 679
Qy 682 KLOINADNIFNR-----HYARVGGANTFNIPGSRBTWTANLRS 721
Db 680 TUSFSIENLDFRDYTTWVGQRAPLIYSPGYGASLYDYKGRGTF--GLNYS 729

RESULT 24
HMUR_YERPE
ID HMUR_YERPE STANDARD; PRT; 676 AA.
AC Q56989;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemin receptor precursor.
GN HMUR OR YP00283.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]_TaxID=632;
RP SEQUENCE FROM N.A.
RC STRAIN=KIM6;
RQ MEDLINE=96386041; PubMed=9026634;

```

```

RA Hornung J.M., Jones H.A., Perry R.D.;
RT "The hmu locus of Yersinia pestis is essential for utilization of
RT free haemin and haem-protein complexes as iron sources.";
RL Mol. Microbiol. 20:725-739(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U60647; AAC64866.1; -.
CC DR EMBL; AJ414141; CAC89146.1; -.
CC DR InterPro; IPR000531; TonB boxC.
CC DR Pfam; PF00593; TonB boxC.1.
CC DR PROSITE; PS00430; TONB DEPENDENT REC.1; 1.
CC DR PROSITE; PS01156; TONB DEPENDENT REC.2; FALSE_NEG.
CC KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 676 HEMIN RECEPTOR.
FT SITE 44 51 TONB BOX.
FT SITE 659 676 TONB C-TERMINAL BOX.
SQ SEQUENCE 676 AA; 74230 MW; 84ED731CB914ACD3 CRC64;

Query Match 5.2%; Score 195; DB 1; Length 676;
Best Local Similarity 19.1%; Pred. No. 7.2e-06;
Matches 147; Conservative 104; Mismatches 306; Indels 214; Gaps 33;

Qy 18 LAALSSSVFAAQADLETVHIKQGR--SYNAIVTEKNGDYSSFAVTVGTKIPASLREIPQ 75
Db 16 LAIACTLPATAQAADTTTQTSSKKHSTDTMTVTATGNERSSF-----EAPM 62
Qy 76 SVSIIITNQOVKDRNVDTPDQLARKTPGLRVLSNDDGRSS---VYARGYSEY--NIDGL 130
Db 63 MVTVIEGNAPTSQTAATAADMLRQVPLTVTGS--GRTNGQDVVMRGYKGQGVLTLDGV 120
Qy 131 PAQMOS--INGTLPNLFADFEVNRGSPGGLFDSGEMGGIV-----NLVRKRPTKA 180
Db 121 RQGTDTGHLNSTFLDPAVLKXIEIVRGPAALLYSGALGGVIAETVDAADMLQPGQNSG 180
Qy 181 PQGHAAGFGTH-----KOYKAEADVSGSLN-----SDGSVGRVMAQTVGASPRPAEK 230
Db 181 YRVYSSAATGDHSGFGLGASAFGRDLDGLSFGTRDIGNIR-----QSNGFN---APND 232
Qy 231 NRHETFYAAADWDINPDVLGAGLYQOORHLAPYNGLPADANNKLPSPQHVFGADVWKN 290
Db 233 ETISNVLAGTWQIDSTQSLSANLRY-----YNNSAIEPKNPQTSAPSTNVMTNRST 285
Qy 291 FKMSHNDVFADLKHVFGNGGKVGKMYRSDRSDSNYAFAGSKLGMK-----TPAGRPG 344
Db 286 IQRDAQ-----LRNLIKPLDQEWLNATAQVYVSEVINEARPG--- 323

```



```
Db 50 VSTIADIRKPNVARDVSKIIITWPGVNLGTNSTSGQGNRRQIDIRGMGPENTLIIID 109
Qy 129 GLPAQMOS-----INGTLNLPFAPDRVEMRGPSGLFDSSGEMGIVNLV 173
Db 110 GKPVSRRSVRQGRGERDTRGDTSWPPM--IERIEVLGRFAAARYGNGAAGGVNII 167
Qy 174 RKRPTKAFQGHAAAGFT--HKQ-----YKAEADVSGSLNSDGSVGRVMAQTVGASPPA 227
Db 168 TKKGSGEWHGSDAYFNAPEHKEGATKRTNPSLTGPLGDEFSPR----- 212
Qy 228 EKNRRHETFYAAD-----WDINPDTVLGAGLYQOHRHAPYNGLPADANNKLPSPQ 281
Db 213 -----LYGNLDKTDQADWDIN-----QGHQSARATYA---TTLPAQREG 249
Qy 282 VFGAD-----WNKFKQNSHDVFDLKHFGNGGKVGKVMRYS--DRDADSNYAFAGSKL 334
Db 250 V-INKDVGVVWD-----FAPLQSELEAGYSRQGNLYAGDTQNTNSDSVTRSKY 299
Qy 335 GSKT-----PAGRPGCNTA---DDKACAVG 356
Db 300 GDETNRLYRQNYALTWNGGWDGVTTSNWVQYEHTRNSRIPEGLAGTGEKFNKATQDF 359
Qy 357 LGTEIKQKALADASYSPFRIGNTANEFVIGADYNRRP---STNEQGRITLYARGGLA 412
Db 360 VIDLDVWMLHSEVNLDPIDFLVQI---LTLTGWNQRMKDLSSNTQALQ----- 408
Qy 413 LNEFRSIPQVDLIANARKGVGRVGSHTVAT--ENLDEFGIYKSTPHPADGLSLGGRLG 470
Db 409 -----TNTGGAIDGVSTDRSPYSKAEIISLFAENNMELTDSITVTPGLRFD 455
Qy 471 HYKIES-----GEGKTLHKASKTKFTGYAGAVYDLNNDNSGLYLSLSQLYTP- 516
Db 456 RHSIVGNWNWSPALNISQGLGDDFTLKMGTARAYK--APSLYQTNPNYILYKSGQCYASA 513
Qy 517 -----QTNLDADGKLLKPRQGNQFVGKGYKGYMDRLNARVSYFMKDKNAAPLNPNK 571
Db 514 GCVYQGGND-----LKAETSINKIEGUE--FKRDGWLAVGWTFNRDYNRKI----- 558
Qy 572 KTRYAALG-----KRVMEGVETEISGAV--TPKQIHAQYSYLHSQIKTA 614
Db 559 EAGYVAVGQNAVGTDLQWQNVKPAVVEGLESLNVPVSETVMWNTNITY-MLKSENKTT 617
Qy 615 SNRRDGGFLMPKHSANLWTTYYQVTPBELTGGGVNAMSGIT-----SSAGMAG 664
Db 618 GDR-----LSIIPEYTLNSTLSQAREDLSM-----QTTFTWYKQOPKKNYKGPAP 666
Qy 665 G-----YATFDMAAYRFTPKLQIADNINRHY-----YARVG 700
Db 667 GPETKEISPYISVIGLSATWDVTQVNSLTGCGVDNLFKRLWRAGNAQTGDLAGANIAGA 726
Qy 701 GANTFNIGSERTANLRYSF 722
Db 727 GAYTNEFG--RTWMSVNTHF 746

RESULT 26
RHTA_RHIME STANDARD; PRT; 746 AA.
AC O923Q5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rhizobactin receptor precursor (TonB-dependent siderophore receptor
rhtA).
GN RHTA OR Ral265 OR Sma2414.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSma (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti.";
RL J. Bacteriol. 193:2576-2595 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Garjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federpiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSma megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888 (2001).
CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AF110737; AAD09419.1; -
DR EMBL; AF007312; AAK65923.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 746 RHIZOBACTIN RECEPTOR.
FT SITE 40 47 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MW; 16AE44A4025D5B23 CRC64;

Query Match 5.1%; Score 193.5; DB 1; Length 746;
Best Local Similarity 18.9%; Pred. No. 1.le-05;
Matches 163; Conservative 116; Mismatches 308; Indels 277; Gaps 38;

Qy 10 INMTAATVLAALSSSVFAAQAD-----LETHIKGQRSYNAIVTEKNGDYSFATV 62
Db 9 ISFCVFWVVGFTGVAQEPANQSEAVTSLEEIVVTGGRSAQOI-----SEIARTI 60
Qy 63 ----GTXIPASLRIPQSVSIITNQOVKDRNVDFD-----QLARKTGLRLVS 107
Db 61 YVVDSDIQAEAR-----SGTKLQOILGETIPSPDAGARTSFQGNLRPLILV- 113
Qy 108 NDDGRSSVYARGYEYSEYNIDGLPAQMQSINGTLNLPFAPDRVEMRGPSGLFDSSGEMG 167
Db 114 --DGVSMNSAR-----SLSRQFDAIDP-----FNTERVEVLGATAIYGGNA-TG 155
Qy 168 GIVNLVRKPTKAFQG-HA--AAGFGTHQYKABADVSGSLNSDGSVGRVMAQTVGASP 224
Db 156 GIINIITKKGDAEPGLHAEVTVGMG-----SGFAGSQDFDRNAAGAVTYNSENWDA 207
Qy 225 RPAENRHEHETFYAADWDINPDTVLGAGLYQOHRHAPYNGLPADANNKLPSPQH--- 281
Db 208 RLSIAGNRGTAFYDGSGLTLLIPDITQTSTAFNERIDLMSGISGYQIDDRRVERFSGPDS 267
Qy 282 -----VFVG-----ADNMKPKONS-----HDVFAD----- 301
Db 268 QDSYDGLYGYPPFAALADPSLFETRSGYESDFNPQTRRSLNVTYDNDVFGQQLLQ 327
```



```

RT common with the RNA p1 promoter family of Escherichia coli." ;
RL J. Bacteriol. 173:1757-1764(1991).
RN [4]
RP SEQUENCE OF 456-614 FROM N.A.
RC STRAIN=RD020;
RA MEDLINE=931069943; PubMed=8093236;
RT Dougherty T.J., Thanassi J.A., Pucci M.J.;
RT "The Escherichia coli mutant requiring D-glutamic acid is the result
RL of mutations in two distinct genetic loci." ;
RN J. Bacteriol. 175:111-116(1993).
RN [5]
RP SEQUENCE OF 21-32.
RC STRAIN=K12 / EMG2;
RA MEDLINE=97443975; PubMed=9296646;
RT Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RL in the genome of Escherichia coli K-12." ;
RN Electrophoresis 18:1259-1313(1997).
RP MUTAGENESIS OF TONB BOX.
RX MEDLINE=90078094; PubMed=2687240;
RA Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C.,
RA Kader R.J.;
RT "Point mutations in a conserved region (TonB box) of Escherichia coli
RT outer membrane protein Bmb affect vitamin B12 transport." ;
RL J. Bacteriol. 171:6526-6533(1989).
CC -1- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR
CC BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10112; AAA23524.1; -
DR EMBL; U00006; AAC43072.1; -
DR EMBL; AE000471; AAC76948.1; -
DR EMBL; M57568; -; NOT ANNOTATED_CDS.
DR EMBL; L14556; AAA23676.1; -
DR PIR; A21892; QRECBT.
DR EcGene; EG10125; btub.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Cobalt transport; Transport; TonB box; Signal;
KW Phage recognition; Receptor; Complete proteome.
KM SIGNAL 1 20
FT CHAIN 1 614 VITAMIN B12 RECEPTOR.
FT SITE 26 33 TONB BOX.
FT SITE 597 614 TONB C-TERMINAL BOX.
FT MUTAGEN 28 28 L->P: INACTIVATE UPTAKE.
FT MUTAGEN 30 30 V->G: INACTIVATE UPTAKE.
FT CONFLICT 162 162 A->G (IN REF. 1).
FT CONFLICT 377 377 A->R (IN REF. 1).
SQ SEQUENCE 614 AA; 68407 MW; AB43CC46A991FF95 CRC64;
Query Match 5.1%; Score 191; DB 1; Length 614;
Best Local Similarity 19.7%; Pred. No. 1.2e-05;
Matches 153; Conservative 109; Mismatches 279; Indels 236; Gaps 36;
Oy 15 ATVALAASSSFAAQTADLETVHIKQGSVAIVTEKNKGDYSSPAVTVGTIPASLREIP 74
Db 5 ASLILTCGVTAFSAQAQD-----TSPTTLV-----VTANRREFP 38
Oy 75 QS-----VSIIITNOQVDRNVDTFDOLARKTPGLRVLSNDG--RSSVYANGVEYS--Y 125
Db 39 RSTVALPTVTVTDIDPWOOSTSVNDVLRRLPGVDITONGSGGOLSSIFIRGTNAASHVLV 98

```



```

Db 612 RGLNLNVEFGEVPGFTRLASATWLDKLTAAASANNNG-----NDAGVANYQ-----L 660
Qy 644 TIGG-----GVNMSGITSSAGMHAG-----GYATPDAMAAYRTPK---LKIQ 684
Db 661 VEGGEVDIPVEGLTATGTGVVRSQGYANBANTLKAKPMTRLDLGVRITMPMDTSLTMR 720
Qy 685 INADINFNHYARVGAGANTFNIPGSERTWTANLRYSF 722
Db 721 ANIENTNERWESVEDSGTYIQGDPRALKLSVMDP 758

RESULT 29
TBPL_NEIMO STANDARD; PRT; 915 AA.
AC 001996;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Blawie G.D., Teat J., Paruchuri D.K.,
RA Thompson S.A., Spurling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
RT utilization and is homologous to TonB-dependent outer membrane
RT receptors."
RL J. Bacteriol. 174:5788-5797(1992).
CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL, M96731, AAA25503.1, -
DR PIR, A43335, A43335.
DR InterPro, IPR000531, TonB_boxC.
DR Pfam, PF00593, TonB_boxC.1.
DR PROSITE, PS00430, TONB_DEPENDENT_REC_1; 1.
DR PROSITE, PS01156, TONB_DEPENDENT_REC_2; 1.
KM Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 898 915 TONB C-TERMINAL BOX.
SQ SEQUENCE 915 AA, 102213 MW, 697CFA810104222F CQC64;

Query Match 4.6%; Score 172; DB 1; Length 915;
Best Local Similarity 21.4%; Pred. No. 0.0042;
Matches 172; Conservative 100; Mismatches 291; Indels 242; Gaps 43;

```

```

Db 101 IRMDNRVSLTVDDL-AQIQSYTAQALGTRTAGSSAINEIEYENKAYAEISGSNS 159
Qy 159 LFPDSEMGQIVNLVYKRPPTKAFQGHAAAG-FCITHQYAEADVSGSLNSDGSVRGRVVA 217
Db 160 VEGGSGALAGSV-----AFQITADVDIGEGRQKGIQSKTATVSGKRGITQSLTALA 210
Qy 218 QTVGASPRAPKNNRHETFEYAAADWDINPDVLAGVLYQQRHLADY----- 264
Db 211 GRIGAEALLIITGRRA-----GEIRAHMAAGR-VQSFRLAVDDGSKAYFIVEE 262
Qy 265 ---NGLPADANNKLPSPQHVRYGADW-----NKRKM---SHDV----- 298
Db 263 ECKNG-----GHEKCKANPKKDVGEDKQTVSTRDYTGPNRFLADLSYESRSLRPPGF 318
Qy 299 -PADLGHYFGNGGYGVKGRYSRDA-----DSNYAFAGSKLAKMPAGR---P 343
Db 319 RENKSHYTG-GILERTQOTFTRDTVPALTKAVFDNMQOAGSLRNGKYAGHHKYG 377
Qy 344 GCNTADDKACAVG--LGTEIKOKALAFDASYSRPFRLGNTANEFYVADYNEFRSTNEOG 401
Db 378 GLFTSGENNAAPVGAEGYGVFVDETHTKSRYGLEYVYTNADK--TWADYARLS----- 429
Qy 402 KTLTVARGLAL-NEFRSIPQVDLIANARKGVRYGSHVATENLDEFG---LYGKSTH 456
Db 430 ---YDRQIGLDNHFO---QTHCSAD---GSDKYCRPADKPFYSYKSDRYVYGBS--- 476
Qy 457 PADGLSLIGGRLGHYKHISGEGKTLHKASKTF-----TYGAGVAYDLNDNN 504
Db 477 -----HKLQAAFPKFSFTAKIRHNLISVNLGYDRGSSNRHOD 514
Qy 505 SYLISLQVLYPTQNLADAKLLKP--ROGNOFEVGYKSGYMDRLANVSYFRMKDKXA 562
Db 515 YYYGSANRAYSLSLTPPQNNKXTSPNGREKNYWSI-----GRNVVTRQCLFGNNT 568
Qy 563 AAPLNPV--NKKTRVAALAKRYVEGYETEISGAVTPKW-QIHAGYSY---LHS---QIK 612
Db 569 YTDCTSRINGKSYAAVADVNLG-----RMADVAGLRLYDRSTHSDGSYS 617
Qy 613 TASN---SDDDGIFLMPGHSANLMTTYQVTEPLTIGGVNMSGITSSAGMHAGYATF 669
Db 618 TGHRTLSNAGIVL--KPADWLDLTYR-----TSTGRLPSPA-- 654
Qy 670 DMAAAYRFTPKLK-LOINDNIFNR 693
Db 655 -EMYGWRSQDKIKAVKIDPEKSPNK 678

RESULT 30
TB12_NEIMB STANDARD; PRT; 908 AA.
ID TB12_NEIMB
AC 006987;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=491;
RN 11
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.
RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;
RX MEDLINE=93345825; PubMed=8344530;
RA Legrain M., Schryvers A.B.;
RA Jacobs E., Mazatrin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
RL Gene 110:73-80(1993).
RN 12
RP SEQUENCE OF 25-45.
RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;
RX MEDLINE=93307625; PubMed=8319886;

```


RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
 RA Borriello S.P., Holland J., Parsons T., Williams P.;
 RT "Antigenic relationships of transferrin-binding proteins from
 RT *Neisseria meningitidis*, *N. gonorrhoeae* and *Haemophilus influenzae*:
 RT cross-reactivity of antibodies to NH₂-terminal peptides.";
 RL FEMS Microbiol. Lett. 109:85-91(1993).
 CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
 CC TRANSFERRIN UTILIZATION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- INDUCTION: BY IRON STARVATION.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z15129; CAA78831.1; -;
 DR PIR; JN0819; JN0819.
 DR PIR; S33154; S33154.
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.
 DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
 DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
 KW Outer membrane; Receptor; Signal; TonB box.
 FT SIGNAL 1 24
 FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.
 FT SITE 38 45 TONB BOX.
 FT SITE 891 908 TONB C-TERMINAL BOX.
 SQ SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;

Query Match 4.4%; Score 167.5; DB 1; Length 908;
 Best Local Similarity 18.2%; Pred. No. 0.00085;
 Matches 180; Conservative 125; Mismatches 330; Indels 355; Gaps 49;

QY 3 QPMVSFRINMTAATVLAALSSVFA-----AQTADLETVHIKQSYNAIVTEKNGDY 55
 DB 2 QOHLFRNLCLSLMTAL--PYAENVQAEQAQKQDITQVAKKQKTRDNEVTG-- 57
 QY 56 SSFVTVGTPKIPASREIPQSVIITNQVKORNDVTFDQARKTPGLRVLSNDDGRSV 115
 DB 58 -----LGLVKSSDTLSKEQVLN-----IRDLTRYDPGIAVBEQGRGSSG 98
 QY 116 YA-RGVEYSEYN--IDGLPAOMOS-----INGTLP-----NLFAPDR---VEVMRGP 156
 DB 99 YSIRGMDKNRVSLTVDGV--SQIQSTQAQALGTRTAGSSGAINIEYENVKAVEISKGS 157
 QY 157 SGLFDPSSGEMGIVN-----
 DB 158 NSSEYNGGALAGSVAFTKTAADIIEGKQWGIQSTAVSGKDHALTQSLAGRSGGAE 217
 QY 172 ---LVKRPTKAPQHAAGFGTH-----KQVK-----AEADVSG 203
 DB 218 ALLIYTKRRGRIHAHKDAGKQVQSFNRLVDLDEKKGQSVRYFIVEECHNGYAACKN 277
 QY 204 SINSDGSV---RGRVMAQTVGASPR---PAE-----KNRHH----- 233
 DB 278 KUKEDASVKDERKTSTQDYTGNSRLLANPLEYGSQSWLFRPGWHLNDRHYVGAVLRTQ 337
 QY 234 ETF-----YAAADWDINP-----DTVLGAGY-----LY 256
 DB 338 QTFEDTRMTVPAYFTSE-DYVPSGLKGLGKYSQNDKAERLFVQGGSTLQIGYGTGVFY 396
 QY 257 QORHLAPYNGLPADANNKLPSPHQHVFGADWNKFKMNSHDVFDLKHVFGNGGKGVKM 316
 DB 397 DERHTKNRYGV-----EYVYNAD-----KDTWAD-----YARLSYDROGI 432
 QY 317 RYSDRDAADSNVAFAGSKLGMKTPACRPGCNTADDKACAVGLGTEIKOKALAFDASYSRPF 376
 DB 433 DLNRLQOHTCHSDGSDKNCRPDGNKPYFYKSDMI-----YEERNLFQAVFKAP 485

QY 377 RLGTANAEFVIGADYNRPRSTNEQGRITTLVARGGLALNEFRSIFQVDLIANAR-----KGV 432
 DB 486 DTAKIRHNLSINLGYDRFKS--QLSHSDYYLQ-----NAVQAYDLITPKKPPFPNGS 535
 QY 433 RYGSHTVA---TENLDEFGIYKSTFHPADGLSLIGG-----RLGHYKIESGEG 479
 DB 536 KDNPIRVISIGKTTVTNTSPICRFGNNTYTDCTPRNIGNGYAAVODNVRLGRW-ADVAG 594
 QY 480 -----KTLHKASKTKFTG-----YAGAV-----VDLNDNNSLYL-----SLSOLYTPQT 518
 DB 595 IRYDRSTHSEDKSVSTGTHRNLWNAGVVLKPTWMDLTYRASTGPRLPSPFAEMYGWRA 654
 QY 519 NLDADGKLLKPRQGNQFEVG--YKG-----SYMDRLNARVSF-YMKDKN-----AA 563
 DB 655 GESLTKLDLKPESKFNREAGIVFKGDFGNLEASYFNNAIRDLIAFGVETRTQNGQTSASG 714
 QY 564 APLPNPKKTRYAA--LGRVMEGVETEISGAVTPKW-----QIHAGYSYLHSGQIKTAS 615
 DB 715 DPGYRNAQNAIRAGINILGKIDWHGV-----WGLPDGLYSTLAYNRKIKVKA- 762
 QY 616 NSRDDGIFL-----LMPK-----HSANLW-----TTYQ-----VTPELTIG 646
 DB 763 DIRADRTFVTSYLPDAVQPSRYVLGLGVDHPDGIWINTMETYSKAKSVDELGSQLLN 822
 QY 647 GGVNAMSGITSSAGHAGGYATFDMAAAYRFTPKLQIADNIFNRHY----- 695
 DB 823 GNANA-----KKAASRRTRPWTVDVSGYINIKKHLTLRAGVYLLNRYVTWENVRTAG 878
 QY 696 -----YARVGGAFTNPISGERTWTANLRY 720
 DB 879 GAVNQHKNVGVYNYRYPAAAPGRNYTFSELMKF 908

RESULT 31
 HXC2_HAEIN
 ID HXC2_HAEIN STANDARD; PRT; 725 AA.
 AC P45357;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Heme/hemopexin utilization protein C precursor.
 GN HXUC.
 OS *Haemophilus influenzae*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC *Haemophilus*.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DL42 / Serotype B;
 RX MEDLINE=95270579; PubMed=7751272;
 RA Cope L.D., Ygeev R., Mueller-Eberhard U., Hansen E.J.;
 RT "A gene cluster involved in the utilization of both free heme and
 RT heme/hemopexin by *Haemophilus influenzae* type b.";
 RL J. Bacteriol. 177:2644-2653(1995).
 CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
 CC CONCENTRATIONS.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE D OF
 CC *H. INFLUENZAE*.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U09840; AAA87059.1; -;
 DR InterPro; IPR000531; TonB_boxC.
 DR Pfam; PF00593; TonB_boxC; 1.

DR PROSITE, PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
 DR PROSITE, PS01156; TONB DEPENDENT REC_2; 1.
 KM Outer membrane; Transport; TonB box; Signal; Receptor.
 FT SIGNAL 1 19 OR 21 (POTENTIAL).
 FT CHAIN 20 725 HEME/HEMOPROTEIN UTILIZATION PROTEIN C.
 SQ SEQUENCE 725 AA; 80838 MW; FC78865020CB5BCE CRC64;

Query Match 4.4%; Score 165.5; DB 1; Length 725;
 Best Local Similarity 20.1%; Pred. No. 0.0085;
 Matches 167; Conservative 115; Mismatches 331; Indels 217; Gaps 43;

QY 5 MGVPRNMTAAVYLAALSSVFAQTADLETHTKQR--SYNAVTEKNQDYSFAVTV 62
 DB 1 MEEKSLIAIATLVYANA--LAQSVELDSIVIAIRDPSPRAYPEKQSK--DSLRSQ 56
 QY 63 GKIPASLEIPQSVSIITNOQVKNRVDTP--FDLARTPELRYLSD-----DGRS 113
 DB 57 ATSVAAALSDIP-----NVDIGGSRSTIAQK--PVRIGLSDNRVVQVIVDVR 101
 QY 114 SVYARGYSEYNIDGLPAQOMOSINGTLNPLFAFDREVMKSGSLFDSGEMGIV--- 170
 DB 102 QNFDLAHRGSYF---LPMNL-----IQIEVIAKGPSSSLMGSALGVVAMR 145
 QY 171 ----NLVR--KRPYKAFQGHAAAGFTHKQY-----KADAVSGSLNSDGVGRV 215
 DB 146 TENALDLKNNDKFQYKIQGYQTANNLSERDVSVEPAANDKFDVLISGFYNADNL---- 201
 QY 216 MAQTGASRPAPKNNRHTFY-----AAADNDIPDTVLGAGYLYQ--RHLPYNGLP 268
 DB 202 -----RTGKGNKLNATYAKQFGGLKFGQINDANRVELSHETRKQTPANN-- 250
 QY 269 ADANNKLPS-----LPOHVEFGADMNFKKNSHDVFDLKHYPNGGKGYKMEYSR 321
 DB 251 -EVENLMEQITDQIREFHKKNNSSPPKAKPSQEEFYSVKTRFSGVS--LTDQQLPQ 308
 QY 322 DADSVAFAFGSKLGMKTPRGRPGCNTADKACAVGLGTEIKOKALAPDASYPFLGNT 381
 DB 309 STVFVYVLTLPNDPYLNTHTALYNNKTIKEQRKVS--GVKDQVTLT-----TRGILRNS 361
 QY 382 AN-----EFYIGADYNFRSTNEGRTTLVARGGLANERPISIQVULINARQVAGYSH 437
 DB 362 SELSHISFYGVDMDKIRTERGTNNKDA-----QFPADY-----NAN-----SN 403
 QY 438 TVATENLDEFYIGKSTFPADGLSLIGGRGLHYKLESQEGTL-----HKASKTK-- 489
 DB 404 TTGVYLIATIPLEGER-----LILSPSVRYHYDTSS--KTIVYKXNNHLSPAIKLTW 453
 QY 490 -----FTYAGAVYDLNDSNLSYLSQLYT-----PQT----- 518
 DB 454 IVTNMLDFTAKYNEAFRAPSMDQERFVSGSHFTSILGRNEINKFVANPNLRPETAOKEI 513
 QY 519 -NLADGLKLRPGNQF--EYGYKSYMDPDLNARVSTYRKDKGAAP-----LNP 568
 DB 514 TANLHEDSLP--KQDGKRIEATYFRNDVKPINKIFINDAKTNNASASAGAGANAP 570
 QY 569 NN-----KTRRYALGRVMEGVETEISGATPKMOIHAGSYLSHQIKTASNRDQIFL 624
 DB 571 NGALLPTKQYQYNTNARLSGIELQ--AQYOTEHLITFTYG-----STGKXKDSGEAL 623
 QY 625 L-MPKHSANILMTTYQ--VTDELITGGVN--ANSGITSSAGMHAGYATFDMAAAYRTP 679
 DB 624 SNIASKIGVGVYALVKDKFTGATVTHYAQRVPKDHSVTPYPIILTLDAAT--AP 661
 QY 680 -----KLTQINADNIENRHYARVAGANTFNI--FGSERTTANLRYSF 722
 DB 682 LKGEWKNLRLDFALENLFDKRYQ-----AFSLMEGTGNNAKISAVYSF 725

RESULT 32
 HSUB NEIMA
 ID HSUB NEIMA STANDARD; PRT; 810 AA.
 AC Q9JWA2;
 DT 16-OCT-2001. (rel. 40, Created)

DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DR HemoGlobin-haptoglobin utilization protein B precursor.
 GN HPUB OR NMA0474.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=22491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parthill J., Achtmann M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
 RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491".
 RL Nature 404:502-506 (2000).
 CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
 HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY
 SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, AL162753; CAB83769.1; --
 DR InterPro; IPR000531; TonB_Box.
 DR Pfam; PF00593; TonB_Box; 1.
 DR PROSITE, PS00430; TONB DEPENDENT REC_1; FALSE_NEG.
 DR PROSITE, PS01156; TONB DEPENDENT REC_2; 1.
 KM Outer membrane; Transport; TonB box; Signal; Receptor;
 KW Complete proteome.
 FT SIGNAL 1 22
 FT CHAIN 23 810
 FT SITE 793 810
 FT SITE 90570 MW; D38DEIDCA3CASA6E CRC64;
 SQ SEQUENCE 810 AA; 90570 MW; D38DEIDCA3CASA6E CRC64;

Query Match 4.3%; Score 162.5; DB 1; Length 810;
 Best Local Similarity 19.2%; Pred. No. 0.0016;
 Matches 168; Conservative 109; Mismatches 306; Indels 293; Gaps 45;

QY 55 YSSFAVTGKTIPLASIREI-----POSVSIITNOQVKNRVDTP-----DQLARTKGLKV 105
 DB 18 FPAFAADPAPOSQOTINELTIVTGTHTOKLGEKIRKTKLVLNDEHDLVRYDGSIV 77
 QY 106 LSNDDRSSVYA--RGVEYS--YNIDGLPAQ-----MOSINGTLPLMFA----- 146
 DB 78 VEGGRAGSNGFTIRGVKRVAINVDGL--AQABRSSEAFQELFGAYGNFANRNTSBE 136
 QY 147 -FDRVEVMKPGSLFDSGEMGIVNLVRK-----RPTKAFQ 183
 DB 137 NFSEVITITGADSLKSGSALGAVVYQTKASDVSEDKPYHLIGTKGSGVKNQSKFS 196
 QY 184 HAAAG-----FGTH--KQYKADAVSGSLNSDGVGRGVMAQTGASRPAP 228
 DB 197 ITAAGRLFGIDLALVYTRFRGKETKRSSTEGDI--EIKNDGV-----YNPLDTG 244
 QY 229 KNNRHETFYAA-----DMDINPDTVLGAGYLYQGRHLAPY-----NGLPADANNKL 276
 DB 245 GPSKYLITVATGVANSQDPQEK--VAKSTLFLKGIYFNQNRIGMIFEDSRIDRFTNELS 303
 QY 277 SLPOHVGADMNKFKNSHDVF-----ADLKHYFGNGGYGVGVGRYSR----- 321


```
Db 304 NLWTGTTTSAATGDYR-HRQDVSYRRRSVEYKNELEHCPWDSLKLRYDKQIDMNTWTW 362
Qy 322 DADSNYAFAGSKLGMKTPAGRPCNNTADDKACAVGLGTEIKOKALAFD-----ASYSRPF 376
Db 363 DIPKNYKRGINGEVVHSPHRIQNTAQ-----WTADFEKQLDFSKAVWAAQYGLGG 414
Qy 377 RLGNTANEFVIGADYNRFRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANARKGVRGYS 436
Db 415 GKGDNAN-----SDYSYF-----AKLYDPKILASNOAKI----- 443
Qy 437 HTVATENLDEFGI-YGKSTFHPADGLSLIGGRLGHYKIESG----- 477
Db 444 -TMLNRSKYFAYNNAFH-----LGGN--DRFRLNAGIRYDKNSSAKDDPKYTT 493
Qy 478 --EGKTLKASKTKETGYA-GAVYDLNNDNSLYLSLSQLYT-----PQT 518
Db 494 AIRGOIPIHLGSEAHAGFSYGVGDFWRFTKHLHL-LAKYSTGFRAPTSDETWLLPFPDPF 552
Qy 519 NUDAGKLLKPRQGNQFVGYKGS-----YM-----DDRNLNARV 552
Db 553 YLKANPN-LKAERKANWELGLAGSKAGNFKLSGFKTKYRDFIELTYMGVSSDDKNNPR- 610
Qy 553 SFYRMKOKNA--AAPLNPNKKTRVAALGKRVMEGVETEISGAVTPKMOIHAG--YSYLH 608
Db 611 -YAPLSDGTALVSSSPWQON--QNRSAAWVGIEFNGTWNLDLSIGLPK-GLHTGLNVSYIK 666
Qy 609 SOIKTASNRDDGIFLLMPKHSANLWTTVQVTPELTIG-----GGVNAMSGITSSA-- 659
Db 667 GK-ATQNGKETPINALSP-----WTAV-----YSLGYDAPSKRWGINAYATRTAAKPK 714
Qy 660 -----GMAAGGYATFDAAVYRFTPKLKLIQINADNI FNRHYA----- 697
Db 715 SDTVHSNDLNNPWPYKHSKAYTLFDLSAYLNIQKQVTLRAAAYNITNKQYTTWESLRS 774
Qy 698 -----RV-----GGANTENIPGSERTWTANLRY 720
Db 775 IREFTGVNRVDNKTTHAGIORFTSPGRSNTTEAKF 810

RESULT 33
ID TB11_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RA (1)
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CCUG 37608 / M982 / Serogroup B / Serotype 9;
RX MEDLINE=93345625; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RA Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
RL Gene 130:73-80(1993).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON STARVATION.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL; Z15130; CAA78833.1; -.
DR PIR; JN0821; JN0821.
DR PIR; S33156; S33156.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 894 911 TONB C-TERMINAL BOX.
SQ SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E6 CRC64;

Query Match 4.3%; Score 161; DB 1; Length 911;
Best Local Similarity 18.3%; Pred. No. 0.0024;
Matches 178; Conservative 133; Mismatches 346; Indels 314; Gaps 49;

Qy 3 QFMSVFRIN--MTAATVLAALSSVFA--AQTADLETVHIKQRSYNAIVTEKNGDYSS 57
Db 2 QOOHLFRNLNLCLSMTALPAYAENVQAQAQEKQLDTIQVKAQKQKTRRDNEVTG---- 57
Qy 58 FAVTVGTKIPASLREIPQSVSIITNOQVKDRNVDTFDOLARKTLCRLVLSNDDGSSVYA 117
Db 58 -----LGLVKVTADTLSKEQVLD-----IRDLTRYDFGIAVVEQGRGASSGYS 100
Qy 118 -RGVEYSEYN-IDGLPAQMOS-----INGTLP-----NLFAFDR---VEVMRGPSPG 158
Db 101 IRGMDKRVSLTVDGL-AIQISYTAQAALGTRTAGSSGAINEIYENVKAVEISKGSNS 159
Qy 159 LFDSSGEGGVIVNLVRKPTKAFQGHAAAG-FGTHQYKAEADVSGSLNSGSGVRGMYA 217
Db 160 VEQSGSALAGSV-----AFQTKADDVIGEGROWGIOSKTAYSGKNRGLTQSIALA 210
Qy 218 QTVGASPR-----PAEKNRHEFP----- 236
Db 211 GRIGGAEALLIHTGRRAGEIRAHEDAGRGVGSFNRNLFVEDESDSYAYFIVEDECEGKNYE 270
Qy 237 -----YAAADWDINPDTVLGAGLYQ-QR 259
Db 271 TCKSKPKDVVGKDERQTVSTRDYTGPNRFLADPLSYESRSLFRP-----GFRFENKR 324
Qy 260 HLAPYNG-----LPA-----DANNK-LPSLPQH----- 281
Db 325 H--YIGGILEHTQTFDTRDMTPVAFLTAKAVFDANSKQAGSLPQNGKYAGNHKYGGLFT 381
Qy 282 -----VFVGADWNKFKMNSHDVFADLKHVFGNGGKVGKVMREY-----SDRADSNYA---F 329
Db 382 NGENGALVGAEGYT-----GVFYDETH-----TKSRYGLEVYVTNADKDTWADYARLSY 430
Qy 330 AGSKLGMKTPAGRPCNNT-ADDKACAVGLGTEIK-----OKALAFDASYSRPPRL 378
Db 431 DRQIGILDNHFQOQTHCSADGSKYCRPSADKPFYSYKSDRVIYGESHRLLOQAPKKSFT 490
Qy 379 GNTANEFVIGADYNRFRSTNEQGRRTLYARGGLALNEFRSIPQVDLIANARKGVRGYSHT 438
Db 491 AKIRHNLNVLNDFRFDPS-NLRHQDYVYQHANRAYSS--KTPPKTANPNDGKS-KPYWVS 546
Qy 439 VATENL--DEFGIYKSTFHPADGLSLIGG-----RLGHYKIESGEG-----KTL 482
Db 547 IGGNVVTGQICLFGNNNTYTDCTPRRSINGKSYAAVRDNVRLGRW-ADVAGLRYDYRST 605
Qy 483 HKASKTKFTG-----YAGAV-----YDLNNDNSLYL-----SLSQLYTPOTNLDADGKL 526
Db 606 HSDGGSVSTGTHRTLSWNAAGIVLKPADWLDLTYRTSTGFRPLSPFAEMYGWGSQVSKAVK 665
Qy 527 LKPRQGNQFVGG--YKG-----SYMD---DRLNARVSFYRMKOKNAAAPLNPN 569
Db 666 IDPKSFKNEAGIVFKGDFGNGLEASWFNNAYRDLIVRGYEAQIKNGKEAKGDPAYLNAQ 725
Qy 570 NKK-TRYAALGKRVMEGVETEISGAVTPK-WQIHAGYSYLH-SQIKTASNSRDDGIFL-- 624
```


Db 726 SAITGGINIGKIDNMGVNDK-----PEGWYTFANRVHVEDIKRADRTDIQSHLFD 780

Qy 625 -LMP-----KHSANLWTTYQVTPBLT--IGGG--VNAMSGITSSAGMHAGG 665

Db 781 AIGPSYVWGLGYDQPEGKGVNGLTYSKAKETELLSRALLNGSRNTKATARTRP 840

Qy 666 YATFDMAAYRFPKLTQINADNIFNRHY-----YARVGANTENIPG 709

Db 841 WYVVDVSGYTTIKKHFPLRAGVYNLNLRYVTWENVRQTAGCAVNOHKNVGVNRYAPG 900

Qy 710 SERTWTANLRY 720

Db 901 RNTFSLEMKF 911

RESULT 34

Y262 HAEIN STANDARD; PRT; 723 AA.

AC P4600;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Probable tonb-dependent receptor HI0262 precursor.

GN HI0262.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RD / KM20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Keitelavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., McElroy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D., Utechtback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M., RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., RA Venter J.C.,

RT "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.";

RT Science 269:496-512(1995).

RL [2]

RN IDENTIFICATION BY MASS SPECTROMETRY.

RX MEDLINE=20137488; PubMed=10675023;

RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B., Gray C., Fountoulakis M.;

RT "Two-dimensional map of the proteome of Haemophilus influenzae.";

CC Electrophoresis 21:411-429(2000).

CC -!- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.

CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).

CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL, U32712; AAC1927.1; -

DR TIGR, HI0262; -

DR InterPro, IPR000531; Tonb_boxC.

DR Pfam, PF00593; Tonb_boxC; 1.

DR PROSITE, PS00430; TONB DEPENDENT REC 1; FALSE_NEG.

DR PROSITE, PS01156; TONB DEPENDENT REC 2; 1.

KW Outer membrane; Receptor; Signal; Tonb box; Complete proteome.

FT SIGNAL 1 21

FT CHAIN 22 723

FT SITE 706 723 TONB C-TERMINAL BOX.

SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FEEA2984 CRC64;

Query Match 4.2%; Score 158.5; DB 1; Length 723;

Best local similarity 19.9%; Pred. No. 0.0026;

Matches 170; Conservative 114; Mismatches 301; Indels 271; Gaps 41;

Qy 5 MSVFRIMTAATVLAALSSVFAAQTADLETTHIKQOR--SYNAITEKNGDYSSPAVTV 62

Db 1 MRPSKLSIATTTTLTVANA---LAQSVELDSINVIAITRDPSPAPVPEKQSK--DSLISKQ 56

Qy 63 GTIIPASLREIPQSVSLITNQVQKDNVDT---FDQARTPELRVLSND-----DGRS 113

Db 57 ATSVADALEDIP-----NVDVGGSSRSIAQK--PNIRGLSDNRVVQVTDGVR 101

Qy 114 SVYARGVEYSEYVNDGLPAQMOSINGTLPLVFAFDRVEVNRGSGGLFDSGEMGATY--- 170

Db 102 QNFDLAHRGSYF---LPMSL-----IQEIEVYKGSSSLSMGGALGAGVAMR 145

Qy 171 ----NLVR---KRPTKAFQGAAGFGTHKQY-----KABADVGSINSQSVRGRV 215

Db 146 TPNALDLKNDKFGVKIRQYQOTANNLSEKQSVFAANDKFDVLISGFYNNADNL----- 201

Qy 216 MAQTVGASPRPAEKNNRHEIFY-----AAADWDINDPTVLGAGYLYQQRHLAPYNGLPA 269

Db 202 -----RTGKNKLNNTAYRQFGGLAKFGQINDANRV-----ELSHRETRFKQTA 246

Qy 270 DANNKLPSPLPQHFVFGADWNKFGNNSHDV-----PADLKHYPNGGYGKV 314

Db 247 PSNNEVENELTNEQITDQIKKFGQKDDLPTPTQSPSPSRSEFYKVKTRLSVST--LT 305

Qy 315 GMRYSRDLADSNYAFAGSKLGKTPAGRCPCNTADKACAVGTEIKOKALAFDASYSR 374

Db 306 DQGIPOGTVFNYVYLPNDPNYALNTALVNNKTIEREQRKVS--GVMDQTKLT-----TR 358

Qy 375 PFLGLGTAN---EPYIGADYVFRSTNEQRTTLYARGLALNERSLPQVDULINARK 430

Db 359 GINLRSSSELSHISFYGVADYMDKIRTRGTNGSDAK-----FRADPY-----NAN-- 405

Qy 431 GVAGVHTVATENLDFEYIGKSTPHPADQLSLGGRLGHYKIESGEGKTL-----HKA 485

Db 406 ----SNTTGVYLIANIPLFEK-----LVSPSVRIDHYTSS---KTVRYKDNHLS 450

Qy 486 SKTKFTG-----YAGAVYDLN-----DNNSLYLSLSQLYTP 516

Db 451 PATKLTWITWNMLDFTAKNENAFRAPSQMERFVSGHFGANTGLDHIRNFVANPVL--RP 509

Qy 517 QT-----NLDADGKLKPRQGNQF--EVGYKSYMDRLNARVSPFMKDKNAAPL 566

Db 510 ETAKNKEITANLHFDRLF--KQGDREKIEATYFRNDVDQFINLKI--FIDAKTSASAGA 564

Qy 567 NPNNKKTTRAAALGKRMGEVETESGAVVPKQOIHAQSYLSHSQIKTASNSRDGIFLLM 626

Db 565 NPNT-----NGALLPK-----NSQYONITNARLSGIELQA 594

Qy 627 PKKS--ANLWTTYQVTPBLTIGGVNAMSGITSSAGMHAGGY-----ATPFAMAA 674

Db 595 QYQTELETLFTNNGSKGDKDQSG--EALSNIAASKIGCVNVALVYDKFTVGATVTHYAA 653

Qy 675 YRFTPK-----LKQINADNIFNRHYARVGANTENIP 707

Db 654 QRRVPPDHVSATYPSYILTLDRATVADLKGEMKRLRDLFALLENLFDKRYQF-----AFSL 707

Qy 708 -PSERTWTANLRYSF 722

Db 708 MEGTGRNAKISAVYSF 723

RESULT 35

HPUB_NEIMC

ID HPUB_NEIMC STANDARD; PRT; 810 AA.

AC P96949;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)


```

RT Escherichia coli K-12."
CC Biochimie 73:1361-1374 (1991).
CC -1- SIMILARITY: TO S-TYPIPHIRIUM ORF NEAR CVSG (AC P25928).
CC CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE GENE CODING FOR
CC THIS PROTEIN IS INTERRUPTED BY A HYBRID IS20/IS90C ELEMENT
CC BETWEEN AMINO ACIDS 839 AND 840.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, AE000237; AAC74483.1; ALT SEQ.
CC EMBL, AE000237; AAC74487.1; ALT SEQ.
CC EMBL, D90778; BAA1880.1; ALT SEQ.
CC EMBL, D90778; BAA1880.1; ALT SEQ.
CC EMBL, D90779; BAA1881.1; ALT SEQ.
CC EMBL, X62680; -; NOT ANNOTATED_CDS.
CC Ecogene; Egl1307; ybda.
CC Hypothetical protein: Complete proteome.
CC CONFLICT 489 I -> V (IN REF. 2).
CC CONFLICT 495 I -> V (IN REF. 2).
CC FT 495 I -> V (IN REF. 2).
CC SQ SEQUENCE 2003 AA; 205949 MW; B83A12C8B53220BE CRC64;

Query Match 3.8%; Score 143; DB 1; Length 2003;
Best Local Similarity 20.9%; Pred. No. 0.12;
Matches 166; Conservative 103; Mismatches 318; Indels 208; Gaps 39;

QY 1 MGFMSVPRINMTATVLAALSSSVPAQAQA-----DLETVHKGORSYNAIVTEKNG 53
DB 626 VDDFSTGVLNGNNNSVTLLAKDLKVVQKATGINSVGDANTVNIQNVLVDRKQADANA 685
QY 54 DYSSRAVYTGKIPALREL--PQSVIITNQVKNRNDTEPQALARKTPGLRYLNDG 111
DB 686 AEFYFPPSGVINGVGDNNVTLDGKLTIVSDSEVTSKOSLFFGSAEKISGLVVG--DG 743
QY 112 RASVYARVGE-YSEVNIIDGLPAQMOSINGTLPFLFADFVREVRMGPSGLF-----DSSGE 165
DB 744 NTVNNMGIELIGERKALADGSGVTS---LRTGYSTSYIVISGSSSVYINDDTTISGE 799
QY 166 ---MGIVNLVARK-----RPTKAFQGHAAAGFGTHK---QYKADAVSGSL 205
DB 800 PFLGAGVIRVODKALLIGSGATLTMDIDSPHH---GRTVIEIQILGFAFVTE- 853
QY 206 NSDGSVRGVMAQTGASPRPA-----EKKNRHETFYAAADMDINDPTVYAGVLYQQRH 260
DB 854 NTTGINSSTISLQNGKDPAPSPVILLATGSGAT-----NAGTITGK---VTEQH 901
QY 261 LAPYNGLPADA-----NNKLPSLPQHVFGADMNKEKNNSHVFPADLKHYRNGGKGV 314
DB 902 -SVFNKYSITGTSNSPIFNNDVSSITGLV---AOSNSTITNTDGIIDL---YGRSGVMIL 954
QY 315 GMRYSRDR-----ADSNYAFAGSKLGKMTPAKPCNTADKCAVGLGTEIKOKALAF 368
DB 955 ALADSTAEKNGKTLTDSMWYDANDTTAMRDIAS---NSAIDRGTEGVGCTD----- 1002
QY 369 DASYSRPFPR-----LQNTANEFTVIGADYRFRSTN---EGGRITLVYARGLALNEFRS 418
DB 1003 ---SYSGAGKAAATVAINQLGVTITYNAGAGAAVAGASNTVINQGTINLEKNGY----- 1053
QY 419 IQQVVDLAAKRGKVRGSHVTAENLDERGIY-----GKSTHPADGLSLIGGGRGLGH 472
DB 1054 ---DDSLAATLVGMVAVEHGTALN--DQGTVININVGTFGAFYNDGTG--TIVVGTITCF 1108
QY 473 KI-ESGE-----GKTLHKASKTKFTGTGAYGVYL---NDNNSLYLSLSQL 513
DB 1109 GVCQSGNEVYNNNDPTSLITYTGDDTITRSGEYTLNKSAAVVDTKLAGNVNNSGTLISGDI 1168
QY 514 YTPQTNLD-----ADGKLLKPRQNGQFEVYGYGSTMDD--RLNAAVSYFRMK 558

```

```

DB 1169 TVSSGILENTSGGIINLVKLDKGAIVIKAGVNTNVVDSGILLNNAAGETAQITM----- 1224
QY 559 DKNAAPLPLN-PNNKKTRYAALGRVMEGVETEISGAVTPKQTHAGYSYLHSQIKTASNS 617
DB 1225 --NAGADSSLVNNTG-----INKIVNAGVFENNSSGVTGRMSAGG-----VFNN 1268
QY 618 RDDGIFLMPKHSANLWTVYVTELTIGGGVNAAGSITSSAGMAGATPDMAAAYRF 677
DB 1269 QDTGALIM-----RGAALTGTAVANNSEGTMLG-----SSSEGNNTG----- 1304
QY 678 TPKLKQIVADNLEN 692
DB 1305 ---MLEVNNNSAFN 1315

RESULT 37
HXCL_HAEMIN STANDARD; PRT; 744 AA.
ID HXCL_HAEMIN
AC P44523;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Heme/hemopexin utilization protein C precursor.
GN HKUC OR H10113.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-I., Glodex A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512 (1995).
CC -1- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CONCENTRATIONS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF
CC H. INFLUENZAE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, U32696; AAC21789.1; -.
CC DR TIGR; H10113; -.
CC DR InterPro; IPR000531; TONB boxC.
CC DR Pfam; PR00593; TONB_boxC_1.
CC DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC KW Outer membrane; Transport; TonB box; Signal; Receptor;
CC KW Complete proteome.
CC FT SIGNAL 1 24 POTENTIAL.
CC FT CHAIN 25 744 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
CC SQ SEQUENCE 744 AA; 85043 MW; 74D94F72E1AEC31 CRC64;

Query Match 3.7%; Score 140.5; DB 1; Length 744;
Best Local Similarity 19.5%; Pred. No. 0.046;

```


Matches 162; Conservative 99; Mismatches 269; Indels 301; Gaps 50;

```
QY 41 QRSYNA-----IVTEKNGDYSSFAVTVGTIKIPASLRREIPQSVSIITNQVQKDRNVTF 93
DB 26 QNSSNSREELPIIWTND-----SNKLPGR-----SVLKQKNEIQXQADNA 69
QY 94 DQARKTGCLRVLSN-DDGRSSVYARGVEYSE---YNTDGLPAQWQSI-NGTL---PNLF 145
DB 70 ANLILPGVNMAGFRPGQGLTINMGMDAEDVRQLDQATKSPKQVQGSIFTEPELL 129
QY 146 AFDRVEVMRGPGLPDSSEMGIGVNLVRKPTKAFQGHAAAG-----FQTH-----KQVK 196
DB 130 R--RTVDKGNYPQYNGGPGFAGTKFKETKDARDPFOENQKIGGLFKYGNNSNNQKYS 187
QY 197 ABADYSG-SLNSD-----GSVRGRVMAQTVGASPRP-----AEKNRHETFYAAADWDIN 245
DB 188 TALVLQNEQKNIDLLFGSVR-----NAGDYKRPDNSKILFSKQKQ-KTGLIKLNWQIS 240
QY 246 PTVVLGAGVLY-----QORHLAPYNGLPADANNKLPSPLOHVFGAD--WKKFKM 293
DB 241 PELLTLSSVYGIHKGWEPFAAKRDILP-----KPSLSDIMRYGTDIAW-KRKL 288
QY 294 NSHDVFAULKHYFGNGGKYGKVMRYSDRDADSNYAFAGSKLGMKTPAGRPGCN----- 346
DB 289 ---VTRDQK-----DENYTLKYNL-----PENNPWNLSTQFSY 320
QY 347 ---TADD---KACAVG---LGTE---IKQALAFDASYSRPRFLNGTANEFVIG----- 388
DB 321 SKTQNDMRPKASSGLVSLGNSQSWITVSDLTDFDINTNSTENIKTTVHELLFGQLWLN 380
QY 389 -----ADYN-----RPRSTNEGRRTLVARGGLALNEFRSIPQVDLIANA 428
DB 381 TRNTLMYDKSVRKADYNGYFQPYMPSGRQYTOAFYLD-----QIKWKNIIFFST 432
QY 429 RGVGVGYSH--TVATENLDFGIGYKSTPHPADGLSLGGRLGHYKTESGSKTLHKAS 486
DB 433 ---GVR-YDHINNIGQKNL-----ALKYNDISAG-----HDYS 461
QY 487 KTKFTG---YAGAVYDLNDDNSLYLSLSQY-----TPQTNLDADGKL 526
DB 462 QKNGYWSYLLGLNDVNHVLSFTNFSKTRAPVIDEQYEQFKQSSVPATSLNLEKEM 521
QY 527 LKPROGNQFVG-----YKGSYMDRLNARVFRMKDKNAAPLNPNN 570
DB 522 I-----NOTRVGGIITLNLHLFOENDAFOFRTTYFYNR--GKNEIFKTRGVNCGVNAADTN 574
QY 571 KKT-----RYAALGRVMGVETE-----ISGAVTPKWQIHAGSYLHISOIKTASNS 617
DB 575 NKVCPKIENYRNLPQYVIQGALEAYYQSYTLFGEIT-----YSYVKGKRTSPRN 626
QY 618 RDDGIFLLMPKHSANLWTTYQVTP-----LTIGGGV-----NAMESG- 654
DB 627 -----PWGKTSWIA-EIPPRKATTALGFNPKYLYTVGWAEFVRRQDRSPLSGD 676
QY 655 -ITSGAGMHAG-GYATFDMAAYRTPKLGK---LQINADNIENRYHVARVG 700
DB 677 PRASSWSLPASRGYSLHNLFLSWS-PAKIKGMNVKITVDNLFNRYNYPVLG 726
```

RESULT 38

```
OAR_MYXA
ID OAR MYXA STANDARD; PRT; 1061 AA.
AC P38370;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE OAR protein precursor.
GN OAR.
OS Myxococcus xanthus.
OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.
OX NCBI_TaxID=34;
RN [1]
```

```
RP SEQUENCE FROM N.A., AND SEQUENCE OF 776-796.
RC STRAIN=D2F1;
RX MEDLINE=93328680; PubMed=8335633;
RA Martinez-Canamero M., Munoz-Dorado J., Farez-Vidal E., Inouye M.,
RA Inouye S.;
RT "Oar", a 115-kilodalton membrane protein required for development of
RT Myxococcus xanthus.";
RL J. Bacteriol. 175:4756-4763 (1993).
CC -1- FUNCTION: REQUIRED FOR CELLULAR ADHESION DURING FRUITING BODY
CC FORMATION.
CC -1- SUBCELLULAR LOCATION: OUTER-MEMBRANE ASSOCIATED.
CC -1- SIMILARITY: LOCAL, TO TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S64103; AAB27614.1; -.
DR PIR; A40609; A40609.
KW Outer membrane; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1061 OAR PROTEIN.
SQ SEQUENCE 1061 AA; 114455 MW; EA8C077296352EF0 CRC64;
```

Query Match 3.6%; Score 137.5; DB 1; Length 1061;

Best Local Similarity 19.0%; Pred. No. 0.12; Indels 361; Gaps 51;

Matches 187; Conservative 116; Mismatches 321;

```
QY 5 MSVFRINNTAATVLA--LSSVFVAAQTADLETV-----SSFAVTGKIPASLRIPQSV- 77
DB 1 MHLNRVLRQETGVVVAAGLLYGSAAFAQSSTIIGTVIDAQSRQPAADVVTATSPNLQGEQ 60
QY 43 SYNAIVTEKNGDY-----SSFAVTGKIPASLRIPQSV- 77
DB 61 T---VVTDAQNGYRIPOLPPGDYTLRPEKEQPKYARSAIQRLNRTIRNVNELLPEALG 117
QY 78 -----SIITNQVQKDRNV-----TFDQARKTPGLRVLS 107
DB 118 EVVEIVGAPPTIDVGSITGMVNVDOEFIKRIAVARPGKGATSPFSLAELAFQAG--- 174
QY 108 NDDGRSSYARYGEYSYNDGL---PAQMOSINGTLPNLFAFDREVEMRGPGLPDS 163
DB 175 NDYGVISINGSTSPENGIVVDGLSTNDPA--FGVNASPLSIEFQVDVNIITG--GYMPEF 230
QY 164 G-EMGGIVNLVRKPTKAFQGHAAAGP-----GTHQYKAEADVSGSLSDGSGVRGVM 216
DB 231 GRSTGGVINAVTRSGSNEFHGVSVPANWTPGTLETRKQIREGTV-----ITGQNG 281
QY 217 AQTVGASPRPAEKNRHETFYAAADWDINPTV-LGAGY---LYQORHLAPYNGLP-DA 271
DB 282 LQNLG-----DFGATLGGPILKDKLWFFAGFAPSPFTRYQHTRLNALRVUDE 328
QY 272 NKLPSLPQHVFGADWNKFKNNSHDVPADLK--HYFGNGGYGKVMRYSDRDADSNYAF 329
DB 329 GNTIKD--ETDFTVAD--AIPGSARKYVADSRTIYMGKLT-----LIQDHNVSFAL 378
QY 330 AGS-----KLGWKTGAG-----RPGCNTADDKACAVGLGTEIKQKALAFDASYS 373
DB 379 NGTPTSTGGLKLSVNPQSGGLPGVLATRPG-----DFGL-TETKANTTSLALYA 428
QY 374 RPFRLGNTANEFVIGADYNFRST-----NEQGRITL--YAR---GGLALNEFR- 417
DB 429 GAF-----ADKKVLDANLGLWFHQTASTLPDGSNLGDRGTGLAGYGRMVYTPTRALTLEA 484
QY 418 -----SIPQVDLIANARKGVGRYS-----HTVATENLDEFGIYKSTF-----H 456
DB 485 LPEQEGACGSGTPEQLV---RSPVTGVGVGPGFMSDQTLDRYOANAKATYLLNALGTH 541
QY 457 -----PADGLSLGGRL-----GHYKIESGEGKTLHKASK-----TKF 490
```



```

Db 542 VKKAGVDVLLSPDQKAYGVGFQEGSGNNGV-AGQGAHVHARRGYQTGPDSAVTF 600
QY 491 TSYA-----GAVYDL-----NDNNSLYLSQUTP----- 516
Db 601 TQVATTTSTTVGGFLDSDMSIANRVTLNIGVRYDVQALYGNGLSLLLGNQMSPTIGAI 660
QY 517 -----QTNLADGKLKPRQG-----NOPE 536
Db 661 VDPFANGRAKVEVNFARVEQVPLNMDRAFPENRISARRSLAEPGQGTATSCDPSSE 720
QY 537 VQKKSQYMDRLNA-----RVSFYMKKKNAAAPLPPNKKTRVYALGKRWMEGVEI 590
Db 721 SQOATCNTDSNLAIAPESSRNVRFYTGTVGGTPVDPDIK-----AQSSDEIVGAEYEV 776
QY 591 SGAVTPKMQIHAGYSYLSHQIKTA--SNRSDG--FFLMPKHS-----ANLWTQY 638
Db 777 LA-----NTRIGASTHCDMNSVIDMSRDGNTYFLGNPDSGPGFEPPTPRYNDVNT 830
QY 639 VTPELLTGGVNAMSGITSA--GMAAGY-----ATPDMAAAYRTPPLKLO 684
Db 831 VYLNRTFADGWLAAQANYTSLRYGNYPGLFRPRTQGLDPIILSDFDL-----IELL 881
QY 685 INADNI--FNKHYARVGAANTFNI 707
Db 882 ENRTGLPDRTHQIKVFGAKEFNI 906

RESULT 39
VITA_VIRCH STANDARD; PRT; 687 AA.
ID VITA_VIRCH STANDARD; PRT; 687 AA.
AC Q00964; Q9J000;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Vibriobactin receptor precursor.
GN VITA OR VC2211.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RX MEDLINE=92276356; PubMed=1317381;
RA Buterton J.R., Stoeber J.A., Payne S.M., Calderwood S.B.;
RT "Cloning, sequencing, and transcriptional regulation of vita, the
RT gene encoding the ferric vibriobactin receptor of Vibrio cholerae.";
RL J. Bacteriol. 174:3729-3738(1992).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Classical Ogawa 395 / ATCC 39541 / Serotype O1;
RA Liao W.J., Choi M.H., Buterton J.R.;
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.B., Hatt D.H., Hickey E.K., Peterson J.D., Unanue L.A.,
RA Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baes S., Qin H., Diego I., Sellers P.,
RA McDonald L., Ueberlack T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -!- FUNCTION: RECEPTOR FOR FERRIC VIBRIOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; AF030977; AAB6828.1; -.
DR EMBL; AE004293; AAF9535.1; -.
DR PIR; A41905; A41905.
DR TIGR; VC2211; -.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
KW Signal; Receptor; TonB box; Outer membrane; Transport; Iron transport;
KW Complete proteome.
FT SIGNAL 1 37
FT CHAIN 38 687 VIBRIOBACTIN RECEPTOR.
SQ SEQUENCE 687 AA; 76413 MW; 2CB4FA9BBAD1AF87 CRC64;

Query Match 3.6%; Score 137; DB 1; Length 687;
Best Local Similarity 18.4%; Pred. No. 0.071;
Matches 127; Conservative 71; Mismatches 247; Indels 244; Gaps 27;

QY 18 LAALSSVFAAQTADLETVHIKGRSYNAIVTEKNGDYSSFAVYTGTKIPASLREIPQSV 77
Db 22 LSAAMMGFTGSAFAAETQNTSNOEQEMFVL-----VVIGEKTORSIYETSASV 70
QY 78 SIITNQVVDNRVDFPDQ-LARKTPG-----LAVLSN--DQGRSS--VYARGVEYSY 125
Db 71 EV-----FDQOTERTPGATEIIDDLQILPVLVDSGSSNNMPTTRIGIDGSGP 117
QY 126 NIDGLPAQMSINGTLF-----NLFAFDREVMARGPSGLFDSGG 164
Db 118 SVGGL--ASTAGNSPLRLMSIDGRSLYSRIAGPRSLMQQVEIYLGQSYIQGN 173
QY 165 EMGGLVNLVRKRPYKAFQGHAAAGFTGHKOYAEADVSGSLNSD 208
Db 174 TSAGAIVMKSNPTHHFESAVVAGIGESDYSQTAGWISAPITODELAFLRSPDQKRDSF 233
QY 209 -----GVRGRVNAQTVGA-----SPRAPKNNR 232
Db 234 VDLAEPAPGDKTIEMNSVRGKLLYEPALDGFKTTTLTSHMSDRGPOTENINAVAGNEA 293
QY 233 -----HETFYAADWDI-----NPDVLGAGLY-----QQRHLAP 263
Db 294 RRPVYETASFTTAMVILHMLDLFTFENNLYVADSYDRYTPNRSRGDNTDQKEPHIEP 353
QY 264 ---YNGLPADANNKLPSPLEQHVFGADMNKFKGNSHDVPADLKHYFGNGYKVGKRRYS 320
Db 354 LRYIALDGSVN-----TLIGARY--YQSGDDWYIDAASAVPMGRTKAKSVF-- 400
QY 321 RADSDYVAFAGS-----KLGKTPAGPGCNTADDKACAVGL----- 357
Db 401 --AEVYVALTPSINVLNLAGRFEREQVKNVSHPRYKLDVDETSSVFLPLDVAVYPVQOQ 458
QY 358 -----GTBIKQALAPDASYSRPRLGNTANFVIGAD--YNPFRSTNEGQRTLYAR 408
Db 459 TYGIRAAKRYNAGAGLAPNSMQFTGFRPYEFOGSIINMYEYTRHRSHSEVLT---- 514
QY 409 GGLALNEFRSIPQVDLIANARKGVRGYSTVATENUDEFGIYK--STFHPADGISLIG 465
Db 515 -NLFPVDFDSMQTQTS-----SGDVIANLDEASTYGAIGSRWATSSLELF- 563
QY 466 GGRLGHYKIESE--GKTLHKASKTKFTGYCAAVYDLDNNSLYLSLSLYTPQTLVLD 523
Db 564 -ANLGLKTEFEKTTGNTLPRAPRMSANVGLVDFG----- 600
QY 524 GKLKPRQGNQF--EVGYKSYMDRLNA 550
Db 601 -----QGFESSNAAYTGSIFSESQNS 622

RESULT 40

```


WAPA_BACSU
 ID WAPA_BACSU STANDARD; PRT; 2334 AA.
 AC Q07833;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR NI7G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of
 RT Bacillus subtilis 168: evidence for processing of the product of a
 RT gene encoding a 258 kDa precursor two-domain ligand-binding
 RT protein.";
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 RT genome containing the hut and wpa loci.";
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 RA Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 RT containing the lic and cel loci, and creation of a 177 kb contig
 RT covering the gnt-sacry region.";
 RL Microbiology 142:3113-3123(1996).
 CC -!- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 CC MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 CC INTO THE MEDIUM.
 CC -!- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS; THE N-TERMINUS, HAS THREE
 CC 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 CC C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 CC MOTIF REPEATED 31 TIMES
 CC -!- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 CC SIMILARITY TO THE REPEAT IN E. COLI Rhs GROUP OF PROTEINS (RhsA-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L05634; AAA22883.1; -
 CC EMBL; D31856; BAA06656.1; -
 CC EMBL; D29985; BAA06260.1; -
 CC EMBL; D83026; BAA11683.1; -
 CC EMBL; Z99124; CAB15959.1; -
 CC PIR; S32920; S32920.
 CC Subtilist; BG10797; wapa.
 CC InterPro; IPR003305; CEM_CenC.
 CC Pfam; PF02018; CEM_4_9; 1.
 CC Cell wall; Repeat; Signal; Complete proteome.
 KW SIGNAL 1 28 OR 32 (POTENTIAL).
 FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
 FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 504 605 1-1.
 FT REPEAT 636 736 1-2.

FT REPEAT 769 869
 FT DOMAIN 1021 2139
 1-3.
 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
 2-1.
 2-2.
 2-3.
 2-4.
 2-5.
 2-6.
 2-7.
 2-8.
 2-9.
 2-10.
 2-11.
 2-12.
 2-13.
 2-14.
 2-15.
 2-16.
 2-17.
 2-18.
 2-19.
 2-20.
 2-21.
 2-22.
 2-23.
 2-24 (APPROXIMATE).
 2-25.
 2-26.
 2-27.
 2-28.
 2-29.
 2-30.
 2-31.
 SQ SEQUENCE 2334 AA; 258329 MW; B75138CCD278BAA3 CRC64;
 Query Match 3.6%; Score 137; DB 1; Length 2334;
 Best Local Similarity 19.6%; Pred. No. 0.4;
 Matches 165; Conservative 93; Mismatches 328; Indels 254; Gaps 39;
 QY 10 INMTAATVLAALSSS-----VFAQTADLETVH--IKGORSYNAIVTEKNGDYSSFA-VT 61
 DB 1487 VSLKAARTSASQAGSVTKQTIVLGOSANDKPYLTLTGMSKASSVKFDEKDYSLQANVT 1546
 QY 62 -----VGTRKIPASLRE-----IPQS-----VSIITNQVKRNVDTFDQLARKTP 101
 DB 1547 YADGSTGIYNAKFPSTQEMNRAAVIPKTKPINKVDISILFQKSATGTVMFDDIRLIEG 1606
 QY 102 GLRVLSNDGRSSVYAR-----GYEYS-EYNIIDGLPAQMOSINGTLPLNLFAPDRVE---- 151
 DB 1607 SLLTKSTYDSNGYVTKBEDEBLGVATSTDYDETGKKTSETDAKGE-KTTYTYDQADQLTN 1665
 QY 152 --VMRPSGLFDSSGEMGGIVNLVRKPTKAFQGHAAAGFGTHKQYKAEADVSGSLNSDG 209
 DB 1666 MTLNNGTSILHSYKEGNEVSKTIRA-----GADQTYKFEYDMVKLVKVT 1711
 QY 210 SVRGVMAQTGYA-----SPRPAEKNNRHETFYAAADWDINPDTVLGAGLYQORHL 261
 DB 1712 DPLGNVLASEYDANSNLTKTISPNGNEVSLVD-----GTDVRVKSXNGTEKYI 1761
 QY 262 APYNGLPADANKPLPSLPQHVFVGADWNKFKNNSHDVADLKHVFGNGG----- 310
 DB 1762 FTY-----DKNGNETSV-----VNKEQNTTKRTFDNNKRLTELTDGSGSQTWTFPSDS 1811
 QY 311 -----YGVKGMRYSDRDADSNYAFAGSKLG-MKTPAGRPGCNTA-- 348
 DB 1812 KLKTFSWTHGQKGTNQFTYKNLQDMIEKOSTSYSDYDENGWVQTFITNGGGTFS 1871
 QY 349 --DDKACAVGL-----GTEIKQKALAFDASYSRPRLGNTANEFVIGA---DYNRF----- 394
 DB 1872 YDERNLVSSLHIGDKNGGDIILTESVEYDANGNR-----TTINSSASGVQVEYKGLNQLV 1926
 QY 395 RSTNEQGTTLTYARGGLALNEFRSIPQVDLIANARKVGRYSHTV-ATENL-----D 445


```

Db 1927 KETHEDGVIEYTDG-----FGRNRTVTITKD-----GSSSTVNASFFIMNQLTKVND 1975
QY 446 EFGIYKSTFHPADG-----SLIGGRLGH-----YKISEGEGKTLKASKTKE 490
Db 1976 EESISYDKNKNRSDGFEYTWDAEDNLTAATYKGEDEKPRATYKDE-KGNRIKTYNGKY 2034
QY 491 TGYAGAVYLDNNSLYLSLSQLYPTPNLADGKLLKPRQNGPEVGYKGSYMDRLNA 550
Db 2035 TNY-----FYDGDLSLVLY-----ETDADNNVTK-----SYTYGDSGQLSTENG 2075
QY 551 RVSFYMKNKMAAALPNPNKKT-----RYAALGRKMEGVETISGAVYPRKQIHAGYISY 606
Db 2076 KRYFHYNAHGDIIAISDSTGKTVAKYQYDAMG-----NPTKTEASDEVXDNRYRYAGYQY 2131
QY 607 LHSQITKASNRDDGIFLMPKH-----SANTLWTTQ 638
Db 2132 -----DEBTGLYIMARYEPNRNGVFLSLDPPSGDSDLDONGYAYGNNPNVKN 2181
QY 639 VTPE-----LTIGGGVNAMSG-----ITSSAGMAGGYATFD-----AMAAVEPTPK 680
Db 2182 YDPDGHVWLVVYNAGPAAVDGYKAYKSGKMGKAWAAMAASNFGPKIFKASAPAYFTTK 2241

```

```

RESULT 41
HGBC_HAEIN STANDARD; PRT; 993 AA.
ID HGBC_HAEIN
AC 09KIV0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin binding protein C precursor.
GN HGBC.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OK NCBI_TaxID=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NTM1 N182;
RX MEDLINE=20316037; PubMed=10858226;
RA Cope L.D., Hxkal Z., Hansen E.J.;
RT "Detection of phase variation in expression of proteins involved in
RT hemoglobin and hemoglobin-haptoglobin binding by nonypeable
RT Haemophilus influenzae.";
RL Infect Immun. 68:4092-4101(2000).
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND
CC IS REQUIRED FOR HEME UPTAKE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -1- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF221060; AAF80178.1;
CC InterPro; IPR000531; TONB_boxC.
CC Pfam; PF00593; TONB_boxC.1;
CC DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1;
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
KW

```

```

KW Receptor; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 993
FT DOMAIN 26 37
FT REPEAT 26 29
FT REPEAT 30 33
FT REPEAT 34 37
FT SITE 47 54
FT SITE 976 993
SQ SEQUENCE 993 AA; 113616 MW; A551BR3BC641612 C64;

```

Query Match 3.6%; Score 136; DB 1; Length 993;
 Best Local Similarity 19.9%; Pred. No. 0.14;
 Matches 129; Conservative 76; Mismatches 211; Indels 232; Gaps 33;

```

QY 5 MSVFRTMTAAVTLALSSSV--FAQTALLETYHIKGRS-----YNAIYENKNGDY 55
Db 1 MAFRLNVLAVSYMGLTGVAAAPTOPTOPTONSNSVSEQLQIVSGSTNSD- 59
QY 56 SSFAVTVGTRIPASLREIPQSVSIITNQO--VKDRNVDTPDLAKRTGLRYL-SNDG 111
Db 60 -----TKTPKIAETVYKTKTLERQANNKD-----IVKYETGVYVEAGRG 103
QY 112 RGSVARGYEYSE--YNIIDLPAMQOSINGTLPNLFAPDRVEVMRQPSGLPSSGEMGI 169
Db 104 QSGFAIRGVDENRVAINIDGL-RQAEITLS-----QGFKELFEGYGNFNT 148
QY 170 VNLVRKPTKAOGHAAAPGTHKQYKAADVSGLNS-----DGSYRGVMQTVASRR 225
Db 149 RN-----GAEIETLKEVNTTKGANSIKSGSGSAGSVYKTKD----- 186
QY 226 PAKNNRHETFYAAADWDINPVTLAGAVLYOORHAPYNGLPADANNKLPSLPQHVFG 285
Db 187 -----ARDYLKND-----YVSYKKGATENNQSPNTLT-----LA 218
QY 286 ADMNKF-----KMSHDVFPADLKHYPNGNGYKGVKMYSDRDADSNYAFASGKLGMT 338
Db 219 GRYKFPDALVWTRRNGHEL-----ENVDYKNADS-----LTGKKR 255
QY 339 PAGRPGCNTADKACAVGLGTEIKQKALAFDASYSRPFRLGNTAN-EFYIGADYRFRST 397
Db 256 EKADP-----YIEQDSTLLKLSFN-----PLENRFTLADLYEHRER 294
QY 398 NEGGRITL-YARGG-----LALNEFRSIPQVDL-----ANARG 431
Db 295 GQDLSTLLKYQSGNETPEVESRHTNDKTKRNNISYSYNSOTPEWDTLKLTYSQRIK 354
QY 432 VR-----GYSHVATENLDEFGIYKSTFHPADGLSLIGGRLGHYKIESGEGKTL 482
Db 355 TPARDEYCDAGVTHQGTEN-----PTGKLTNG-----KITRNGTDL 394
QY 483 HKASKTFPGYGA-VYDIN-----DNNSLYLSLSQLYPTPNLDA-----DGKILKR 530
Db 395 Q--FFEKGTAAAGTYKYPDFTIDTNNQEIYKLTNIGVANDTWYDCLPDKDKGTAKVF 452
QY 531 QGNQPEVGYKGSYMDRLNARY-----SFYRMDKMAAALPNPNKKT 574
Db 453 EGNA-TYGIIGKRESELETEVLNKRKFARIYDK-----DKNNRNR 494

```

```

RESULT 42
PAT2_HUMAN STANDARD; PRT; 4349 AA.
ID PAT2_HUMAN
AC 09NY08; 075091; 09NSR7;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Procadherin Fat 2 precursor (hfat2) (multiple epidermal growth
DE factor-like domains 1).
GN FAT2 OR MEGF1 OR CDH8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
KW

```


OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20202599; PubMed=10716726;
 RA Wu O., Maniatis T.;
 RT "Large exons encoding multiple ectodomains are a characteristic
 feature of protocadherin genes."; Proc. Natl. Acad. Sci. U.S.A. 97:3124-3129(2000).
 RN [2]
 RP SEQUENCE OF 3777-4349 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=98360089; PubMed=9893030;
 RA Nakayama M., Nakajima D., Nagase T., Nomura N., Seki N., Ohara O.;
 RT "Identification of high-molecular-weight proteins with multiple
 EGF-like motifs by motif-trap screening."; Genomics 51:27-34(1998).
 RN [3]
 RP SEQUENCE OF 4142-4349 FROM N.A.
 RC TISSUE=restis;
 RA Pouetka A., Wellenreuther R., Mewes H.-W., Weil B., Wiemann S.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: CONTAINS 32 CADHERIN DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF231022; BAA61928.1; -;
 DR EMBL; AB011535; BAA32463.1; -;
 DR EMBL; AL157443; CAB75663.1; -;
 DR Genew; HGNC:3596; FAT2.
 DR MIM; 604269; -;
 DR HSP; P00740; 1EDM.
 DR InterPro; IPR002126; Cadherin.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001791; Laminin G.
 DR Pfam; PF00028; cadherin; 33.
 DR Pfam; PF00008; EGF; 2.
 DR PRINTS; PR00205; CADHERIN.
 DR SMART; SM00112; CA; 32.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00282; LamG; 1.
 DR PROSITE; PS00232; CADHERIN_1; 14.
 DR PROSITE; PS0268; CADHERIN_2; 32.
 DR PROSITE; PS00022; EGF 1; 2.
 DR PROSITE; PS01186; EGF 2; 1.
 DR PROSITE; PS50025; LAM G DOMAIN; 1.
 KW Transmembrane, Glycoprotein; Repeat; EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 4349
 FT DOMAIN 19 4048
 FT TRANSMEM 4049 4069
 FT DOMAIN 4070 4349
 FT CYTOPLASMIC (POTENTIAL).
 FT CADHERIN 1.
 FT CADHERIN 2.
 FT CADHERIN 3.
 FT CADHERIN 4.
 FT CADHERIN 5.
 FT CADHERIN 6.
 FT CADHERIN 7.
 FT CADHERIN 8.
 FT CADHERIN 9.
 FT CADHERIN 10.
 FT CADHERIN 11.
 FT CADHERIN 12.
 FT CADHERIN 13.
 FT CADHERIN 14.
 FT 1556 1660

FT DOMAIN 1561 1758
 FT DOMAIN 1759 1872
 FT DOMAIN 1969 2070
 FT DOMAIN 2071 2171
 FT DOMAIN 2172 2272
 FT DOMAIN 2273 2372
 FT DOMAIN 2380 2481
 FT DOMAIN 2482 2585
 FT DOMAIN 2586 2691
 FT DOMAIN 2692 2797
 FT DOMAIN 2798 2906
 FT DOMAIN 2907 3011
 FT DOMAIN 3012 3113
 FT DOMAIN 3114 3218
 FT DOMAIN 3219 3321
 FT DOMAIN 3322 3426
 FT DOMAIN 3427 3531
 FT DOMAIN 3532 3642
 FT DOMAIN 3643 3744
 FT DOMAIN 3745 3844
 FT DOMAIN 3845 3944
 FT DOMAIN 3945 4022
 FT DOMAIN 4023 4117
 FT CARBOHYD 39 39
 FT CARBOHYD 210 210
 FT CARBOHYD 280 280
 FT CARBOHYD 330 330
 FT CARBOHYD 459 459
 FT CARBOHYD 568 568
 FT CARBOHYD 627 627
 FT CARBOHYD 655 655
 FT CARBOHYD 789 789
 FT CARBOHYD 996 996
 FT CARBOHYD 1175 1175
 FT CARBOHYD 1303 1303
 FT CARBOHYD 1383 1383
 FT CARBOHYD 1417 1417
 FT CARBOHYD 1904 1904
 FT CARBOHYD 1998 1998
 FT CARBOHYD 2007 2007
 FT CARBOHYD 2165 2165
 FT CARBOHYD 2183 2183
 FT CARBOHYD 2325 2325
 FT CARBOHYD 2368 2368
 FT CARBOHYD 2387 2387
 FT CARBOHYD 2430 2430
 FT CARBOHYD 2470 2470
 FT CARBOHYD 2547 2547
 FT CARBOHYD 2597 2597
 FT CARBOHYD 2654 2654
 FT CARBOHYD 3125 3125
 FT CARBOHYD 3276 3276
 FT CARBOHYD 3310 3310
 FT CARBOHYD 3430 3430
 FT CARBOHYD 3471 3471
 FT CARBOHYD 3601 3601
 FT CARBOHYD 3772 3772
 FT CARBOHYD 3813 3813
 FT CARBOHYD 3840 3840
 FT CARBOHYD 3873 3873
 FT CARBOHYD 3904 3904
 FT CARBOHYD 3989 3989
 FT CARBOHYD 4117 4117
 FT CONFLICT 4160 4160
 FT CONFLICT 4160 4160
 SQ SEQUENCE 4349 AA; 479383 MW; 6C050CA17AD398ED CRC64;

Query Match 3.5%; Score 130.5; DB 1; Length 4349;
 Best Local Similarity 21.9%; Pred. No. 2.7;
 Matches 97; Conservative 52; Mismatches 140; Indels 153; Gaps 21;

OY 358 GTEIKQKALFADASYSRPFLGNTANEFVIGADYNRFR--STNEQGRITLVARGGL---- 411

DB 3234 GTEVLQL-----ATLTP-----GAEKGYRVSGNEQGRFLDARTGILYN 3276


```

QY 412 ALNEPESIPQVLD-IMARKGVNGYSH-----TVATENLDEF--GIYKSTFHPA-- 458
DB 3277 ASLDEFSTPKYFLSTIECSRKSSSLDVTVMVNITDVNEHRQFPQDPYSTRELENALV 3336
QY 459 -----DG-----LSLIGGRGLGHYKIESGEG-----KTLHKASKTKF----- 490
DB 3337 GGVILTVSATDEBGLPNSDITLISLIGNQUGHTTHPKKELQVAKALDREQASISLKL 3396
QY 491 -----TGYAGAVYDLNNSLSYLSQUTPOTNLDAQKLIK-----P 529
DB 3397 RATSQGPPLHEDTDIAIQADVNDNPPRFQALNSTTVGENSPISKVLQLLSDDSP 3456
QY 530 RQGNQEVGVKGSYMDRLNARVSFRMKDKNAAA-PLNPNKKTRYAALGKVMGCVET 568
DB 3457 ENGPPY-----SFRITKGNNGSAFRVPPDGMVLYAEGLSRRAQEWYQL 3499
QY 589 EISGA---VTPKQIHAGYSYLHSQIKTASNSDDGFIILMPK-----HGANL- 633
DB 3500 QIQASDSGIPPLSSLTYSVRVHTVQSHYAPSLPLEIFITVGEDEFQGNVGIHATDRD 3559
QY 634 ---WTVYQVTPBELTIG-----GVNAMSGL-----TSAGM----- 661
DB 3560 PQDTLTVSLABEETLGRHPSVGAHPDKIILAAQGLPRGHVSFNTVSDGCTTTAGVHVY 3619
QY 662 -HAGVATFDA--MAAYRFTPK 680
DB 3620 WHYGDEALQOAMMMGFYQLTPE 3641

```

RESULT 43

```

CHIT1_BACCI STANDARD; PRT; 699 AA.
ID CHIT1_BACCI
AC P20533;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Chitinase A1 precursor (EC 3.2.1.14).
GN CHIT1.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxId=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WL-12;
RX MEDLINE=90368776; PubMed=2203782;
RA Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;
RT "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed
RT its evolutionary relationship to Serratia chitinase and to the type
RT III homology units of fibronectin."
RL J. Biol. Chem. 265:15659-15665(1990).
RN [2]
RP MUTAGENESIS.
RC STRAIN=WL-12;
RX MEDLINE=93366760; PubMed=8103047;
RA Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.,
RA Uchida M., Tanaka H.;
RT "Identification of glutamic acid 204 and aspartic acid 200 in
RT chitinase A1 of Bacillus circulans WL-12 as essential residues for
RT chitinase activity."
RL J. Biol. Chem. 268:18567-18572(1993).
RN [3]
RP CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
RN acetyl-D-glucosamine polymers of chitin.
RN [4]
RP SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
RN [5]
RP SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
RN HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC -----
DR EMBL; M57601; AAA81528.1; .
DR PIR; A38368; A38368.
DR HSSP; P02751; IFNA.
DR InterPro; IPR003610; CEM_5_12.
DR InterPro; IPR001579; Chitinase_18/2.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003962; FNIII_repeat.
DR InterPro; IPR01223; Glyco_hydro_18.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR Pfam; PF02839; CEM_5_12; 1.
DR PRINTS; PR00014; FNTYPEIII.
DR ProDom; PD000471; Glyco_hydro_18; 2.
DR SMART; SM00495; CatB3; 1.
DR SMART; SM00060; FN3; 2.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW Hydrolyase; Glycosidase; Chitin degradation; Signal; Repeat.
FT SIGNAL 1 41
FT CHAIN 42 699
FT DOMAIN 42 460 CATALYTIC.
FT DOMAIN 465 549 FIBRONECTIN TYPE-III (R-1).
FT DOMAIN 560 644 FIBRONECTIN TYPE-III (R-2).
FT ACT_SITE 204 204 PROTON DONOR (PROBABLE).
FT ACT_SITE 200 200 D->N: DECREASE IN ACTIVITY.
FT MUTAGEN 200 200 D->E: NO CHANGE IN ACTIVITY.
FT MUTAGEN 204 204 E->D,Q: LOSS OF ACTIVITY.
SQ SEQUENCE 699 AA; 73677 MW; AC7C9B2E2987643 CRC64;

```

Query Match 3.4%; Score 129; DB 1; Length 699;
 Best Local Similarity 21.1%; Pred. No. 0.26;
 Matches 127; Conservative 47; Mismatches 208; Indels 220; Gaps 27;

```

QY 198 EADVSGSLNSDSVRRWAAQTVGASPRPAEKNNR-----HETVYAADMDINDTVL 250
DB 204 EYFVSGLD-----GNSKRPEDKQNTYLLSKREKLDAAQAVD----- 242
QY 251 GAGYLYQOHLAPYNGLPAD--ANNKLPSLPQHVFGADMNKFNKSHVFDLKHYPGN 308
DB 243 GKLYL-----LTIASGASTVYAAETELAKIAIV-----DW-----INIMYDYN 282
QY 309 GGYGVKMYRSDRDADSNYAFAGSKLGMKTPAGRPGCNTADKCAVGLGTEIKQALAF 368
DB 283 GAWOKI-----SAHNAPLNYPDAAS-----AAGVPDANFNVAAGQG-----HL 322
QY 369 DASYSRFLGNTANSEFVGADYNNRPSRINBOGRTLLVARGGLALNEPESIPQVDLIANA 428
DB 323 DA-----GVPAALVLGVPP-----YGRGW-----DGCQAQ 348
QY 429 RKGVRGSHVTATENLDEFGIYKSTFHPADGLSLIGGRGLGHYKIESGEGKTLHKASKT 488
DB 349 -----GNGYQTCGTGSSVGTWEASFPDYDLEANTINNGYT 386
QY 489 KFTGVAGAV-YDLNNSLSYLSQUTPOTNLDAQKLIKPRQGNQFVGVKGSYMDR 547
DB 387 RYWNDAKPYLYLNASNKRFISYD-----DAE-----SVGYTAVIKSK 425
QY 548 LNAKVSFYMK-DKN-----AAAPLNPNNKT----- 573
DB 426 GLGAMFELSGDRNKTLLONKLAADLPFGTVPVPTTAPSVFGNARSQVTANSVTLAM 485
QY 574 -----RYALGKRVNGG--VETEISGAVTPKQIHHGYSYLH-----SQIK 612
DB 486 NASTDNVGVGTGVNVNNGALASVITGTATISGLTAGTSTYTPFIKAKDAAGNLASAANAV 545
QY 613 TASNRRDGI FILMPKHSANLWTVYQVTPBELTIG-----GVNAMSGLTSAGNH 662
DB 546 TVSTTAQPGDQAQAPAPFNALASTAGTSSITLSWTASTDNVGVTSYDVGNG-TALATTV 604
QY 663 AGGYATFDMAA---YRFPKLTQINAMNINRHHYAVAGGANTFNIGSERTWTANLR 719
DB 605 TGTATISGLAADTSTYTFVAKK--DAAGNVAASNAVSVKTALETNTNGVS-AMQVNTA 661

```


Search completed: December 18, 2002, 06:42:58
Job time : 29.3071 secs